

## Durham Research Online

---

### Deposited in DRO:

18 November 2020

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Ross, Ashley J. and Bautista, Julian and Tojeiro, Rita and Alam, Shadab and Bailey, Stephen and Burtin, Etienne and Comparat, Johan and Dawson, Kyle S and deMattia, Arnaud and duMasdesBourboux, Hélión and Gil-Marín, Héctor and Hou, Jiamin and Kong, Hui and Lyke, Brad W and Mohammad, Faizan G. and Moustakas, John and Mueller, Eva-Maria and Myers, Adam D. and Percival, Will J. and Raichoor, Anand and Rezaie, Mehdi and Seo, Hee-Jong and Smith, Alex and Tinker, Jeremy L. and Zarrouk, Pauline and Zhao, Cheng and Zhao, Gong-Bo and Bizyaev, Dmitry and Brinkmann, Jonathan and Brownstein, Joel R and Rosell, Aurelio Carnero and Chabanier, Solène and Choi, Peter D and Chuang, Chia-Hsun and Cruz-Gonzalez, Irene and delaMacorra, Axel and delaTorre, Sylvain and Escoffier, Stephanie and Fromenteau, Sebastien and Higley, Alexandra and Jullo, Eric and Kneib, Jean-Paul and McLane, Jacob N and Muñoz-Gutiérrez, Andrea and Neveux, Richard and Newman, Jeffrey A and Nitschelm, Christian and Palanque-Delabrouille, Nathalie and Paviot, Romain and Pullen, Anthony R and Rossi, Graziano and Ruhlmann-Kleider, Vanina and Schneider, Donald P and Magaña, Mariana Vargas and Vivek, M and Zhang, Yucheng (2020) 'The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey : large-scale structure catalogues for cosmological analysis.', *Monthly notices of the Royal Astronomical Society.*, 498 (2). pp. 2354-2371.

### Further information on publisher's website:

<https://doi.org/10.1093/mnras/staa2416>

### Publisher's copyright statement:

This article has been accepted for publication in *Monthly notices of the Royal Astronomical Society*. ©: 2020 The Author(s). Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

### Additional information:

### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale structure catalogues for cosmological analysis

Ashley J. Ross<sup>1</sup>,<sup>★</sup> Julian Bautista<sup>2</sup>, Rita Tojeiro<sup>3</sup>, Shadab Alam<sup>4</sup>, Stephen Bailey<sup>5</sup>, Etienne Burtin<sup>6</sup>, Johan Comparat<sup>7</sup>, Kyle S. Dawson<sup>8</sup>, Arnaud de Mattia<sup>6</sup>, Héliion du Mas des Bourboux<sup>8</sup>, Héctor Gil-Marín<sup>9,10</sup>, Jiamin Hou<sup>7</sup>, Hui Kong<sup>1</sup>, Brad W. Lyke<sup>11</sup>, Faizan G. Mohammad<sup>12,13</sup>, John Moustakas<sup>14</sup>, Eva-Maria Mueller<sup>2</sup>, Adam D. Myers<sup>11</sup>, Will J. Percival<sup>12,13,15</sup>, Anand Raichoor<sup>16</sup>, Mehdi Rezaie<sup>17</sup>, Hee-Jong Seo<sup>17</sup>, Alex Smith<sup>6</sup>, Jeremy L. Tinker<sup>18</sup>, Pauline Zarrouk<sup>6,19</sup>, Cheng Zhao<sup>16</sup>, Gong-Bo Zhao<sup>20,21</sup>, Dmitry Bizyaev<sup>22</sup>, Jonathan Brinkmann<sup>22</sup>, Joel R. Brownstein<sup>8</sup>, Aurelio Carnero Rosell<sup>23</sup>, Solène Chabanier<sup>6</sup>, Peter D. Choi<sup>24</sup>, Chia-Hsun Chuang<sup>25</sup>, Irene Cruz-Gonzalez<sup>26</sup>, Axel de la Macorra<sup>26</sup>, Sylvain de la Torre<sup>27</sup>, Stephanie Escoffier<sup>28</sup>, Sebastien Fromenteau<sup>29</sup>, Alexandra Higley<sup>11</sup>, Eric Jullo<sup>27</sup>, Jean-Paul Kneib<sup>16</sup>, Jacob N. McLane<sup>11</sup>, Andrea Muñoz-Gutiérrez<sup>26</sup>, Richard Neveux<sup>6</sup>, Jeffrey A. Newman<sup>30</sup>, Christian Nitschelm<sup>31</sup>, Nathalie Palanque-Delabrouille<sup>6</sup>, Romain Paviot<sup>27</sup>, Anthony R. Pullen<sup>18,32</sup>, Graziano Rossi<sup>24</sup>, Vanina Ruhlmann-Kleider<sup>6</sup>, Donald P. Schneider<sup>33,34</sup>, Mariana Vargas Magaña<sup>26</sup>, M. Vivek<sup>33,35</sup> and Yucheng Zhang<sup>18</sup>

*Affiliations are listed at the end of the paper*

Accepted 2020 August 11. Received 2020 August 10; in original form 2020 May 12

## ABSTRACT

We present large-scale structure catalogues from the completed extended Baryon Oscillation Spectroscopic Survey (eBOSS). Derived from Sloan Digital Sky Survey (SDSS) IV Data Release 16 (DR16), these catalogues provide the data samples, corrected for observational systematics, and random positions sampling the survey selection function. Combined, they allow large-scale clustering measurements suitable for testing cosmological models. We describe the methods used to create these catalogues for the eBOSS DR16 Luminous Red Galaxy (LRG) and Quasar samples. The quasar catalogue contains 343 708 redshifts with  $0.8 < z < 2.2$  over  $4808 \text{ deg}^2$ . We combine 174 816 eBOSS LRG redshifts over  $4242 \text{ deg}^2$  in the redshift interval  $0.6 < z < 1.0$  with SDSS-III BOSS LRGs in the same redshift range to produce a combined sample of 377 458 galaxy redshifts distributed over  $9493 \text{ deg}^2$ . Improved algorithms for estimating redshifts allow that 98 per cent of LRG observations result in a successful redshift, with less than one per cent catastrophic failures ( $\Delta z > 1000 \text{ km s}^{-1}$ ). For quasars, these rates are 95 and 2 per cent (with  $\Delta z > 3000 \text{ km s}^{-1}$ ). We apply corrections for trends between the number densities of our samples and the properties of the imaging and spectroscopic data. For example, the quasar catalogue obtains a  $\chi^2/\text{DoF} = 776/10$  for a null test against imaging depth before corrections and a  $\chi^2/\text{DoF} = 6/8$  after. The catalogues, combined with careful consideration of the details of their construction found here-in, allow companion papers to present cosmological results with negligible impact from observational systematic uncertainties.

**Key words:** catalogues – cosmology: observations.

## 1 INTRODUCTION

The Sloan Digital Sky Surveys (SDSS) began in 1998. Since then, through phases I and II (York et al. 2000), III (Eisenstein et al. 2011), and IV (Blanton et al. 2017), they have used the Sloan telescope (Gunn et al. 2006) in order to amass 2.6 million spectra of galaxies and quasars (Ahumada et al. 2020). The primary purpose of these

observations that simultaneously place a single fibre on hundreds of extragalactic objects has been to create 3D maps of the structure of the Universe. From these maps, we observe the large-scale structure (LSS) of the Universe and thereby infer its bulk contents, dynamics, and structure formation history.

During SDSS I and II, the measurement of the location of the baryon acoustic oscillation (BAO) feature in these maps was realized and developed as a robust and powerful method for obtaining geometrical measurements of the expansion history of the Universe and thus dark energy (Eisenstein & Hu 1998; Cole et al. 2005;

\* E-mail: ashley.jacob.ross@gmail.com

Eisenstein et al. 2005; Percival et al. 2010). This motivated the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) and extended BOSS (eBOSS; Dawson et al. 2016) programs of SDSS-III and SDSS-IV. During these programs, considerable research was completed in order to use the signature of large-scale redshift-space distortions (RSD; Kaiser 1987) in the maps as a robust measure of the rate of structure formation (see Alam et al. 2017 and Alam et al. 2020a for summaries of the developments), thereby allowing dynamical tests of dark energy and general relativity.

However, in order to confidently use these maps for these high-precision cosmological purposes, we must understand and account for how the survey design and operation (including all instrumental effects) impact the structure that we record. In essence, at every location in the observed space (angles and redshifts), we estimate the expected mean density (in the absence of any fluctuations due to clustering). This is commonly referred to as the survey ‘selection’ or ‘window’ function. It can be Poisson sampled by a set of random positions in the observed space (defined by the survey design and performance). Variations in the survey selection function can equally be accounted for by applying weights to either the data or random catalogues.

The complexity (in level of detail) in producing these matched data and random catalogues typically leads to independent public releases as SDSS value-added LSS catalogue products, with publications describing their creation. For SDSS I and II, the details are in Blanton et al. (2005).<sup>1</sup> For BOSS in SDSS-III, the details of catalogues extending to  $z < 0.75$  are in Reid et al. (2016). SDSS-IV eBOSS completed on 2019 March 1, and obtained four distinct samples for studies of large-scale clustering. Here, we describe the details of the creation of LSS catalogues for eBOSS quasars and luminous red galaxies (LRGs). Emission line galaxy (ELG) catalogues are described in Raichoor et al. (2020) and the Lyman  $\alpha$  forest analysis of high redshift quasars is described in du Mas des Bourboux et al. (2020).

The observed data and random catalogues we produce serve the primary purpose of obtaining BAO and RSD measurements from two-point statistics; i.e. the correlation function in configuration space and the power spectrum in Fourier space. The catalogues are, at their highest level, simply tables with one column for each of the 3D and extra columns that account for selection effects or provide weights that optimize these BAO and RSD analyses. The format is meant to allow efficient application of common correlation function and power spectrum estimators. While created to serve this particular purpose, the catalogues are documented and made public<sup>2</sup> in the hope that they will be useful for any LSS study.

For eBOSS, we calculated the catalogues using a development of the MKSAMPLE code, which traces its roots back to BOSS Data Release 9 (Anderson et al. 2012), and is described in detail in Reid et al. (2016). In essence, MKSAMPLE was a framework for dealing with the particularities of the SDSS geometry, data model, and observing strategy. Very few of the original lines of code, or even algorithms themselves, are still used in the final eBOSS MKALLSAMPLES package. However, the underlying philosophy and basic set of necessary tasks remain almost the same. In this paper, we detail the changes and additions in the eBOSS process and describe the final catalogues that are produced.

This paper is part of a series of papers presenting the completed eBOSS DR16 data set and cosmological results derived from it, which are summarized in Alam et al. (2020a). The DR16 spectral reductions described in Ahumada et al. (2020) and the DR16 quasar catalogue produced by Lyke (2020) are vital inputs to the LSS catalogues we create. The LSS catalogues themselves were developed in close collaboration with the studies that obtain BAO and RSD results from the eBOSS DR16 data. For the LRGs, the correlation function is presented and used to measure BAO and RSD in Bautista et al. (2020), and the power spectrum in Gil-Marín et al. (2020). Rossi et al. (2020) presents the analysis of mock catalogues, designed to find and quantify any modelling systematic errors associated with the analysis of these data. The equivalent analyses of the quasar sample are presented in Hou et al. (2020), Neveux et al. (2020), and Smith et al. (2020). The ELG catalogues are presented and analysed in Raichoor et al. (2020), further analysed in Tamone et al. (2020), de Mattia et al. (2020), and supported by simulations of the data presented in Alam et al. (2020b) and Lin (2020). Multitracer analysis utilizing the overlapping volume between the LRG and ELG samples is presented in Wang et al. (2020) and Zhao et al. (2020b). The creation of approximate mocks to be used for covariance matrix estimation for all LSS samples is described in Zhao et al. (2020a). Finally the DR16 Lyman  $\alpha$  sample is presented and analysed in du Mas des Bourboux et al. (2020). A summary of all SDSS BAO and RSD measurements with accompanying legacy figures can be found here: <https://sdss.org/science/final-bao-and-rsd-measurements/>. The full cosmological interpretation of these measurements can be found here: <https://sdss.org/science/cosmology-results-from-eboss/>.

The types of target (quasar, LRG, ELG) are described in Section 2, and the targeting criteria for each summarized. In Section 3, we describe the eBOSS observing strategy. The method used to measure redshifts is summarized in Section 4, and the catalogue creation in Section 5. This section also includes details of how we have corrected for many observational effects including varying completeness, collision priority, close pairs, redshift failures, and systematic problems with the imaging data. This section ends with a review of the statistics for each sample, and provides details of how to use these catalogues. A summary of the work is provided in Section 6.

## 2 EBOSS TARGETS

eBOSS was designed to acquire redshifts for three types of tracers: quasars, LRGs, and ELGs. Each object selected from imaging data for follow-up spectroscopy is an eBOSS ‘target’. The selection criteria and motivation for these target samples are detailed elsewhere. Here, we record the essential details.

### 2.1 Quasars and LRGs

LSS quasar<sup>3</sup> and LRG targets were selected using the same optical and infrared imaging data sets over the full SDSS imaging area.

The optical data were obtained during the SDSS-I/II (York et al. 2000), and III (Eisenstein et al. 2011) surveys using a drift-scanning mosaic CCD camera (Gunn et al. 1998) on the 2.5-metre Sloan Telescope (Gunn et al. 2006) at the Apache Point Observatory in New Mexico, USA. The five-passband ( $u, g, r, i, z$ ; Fukugita et al. 1996; Smith et al. 2002; Doi et al. 2010) photometry was recalibrated by Schlafly et al. (2012), who applied the ‘uber-calibration’ technique

<sup>1</sup>The updated details for samples through DR7 are available at <http://sdss.physics.nyu.edu/vagc/>.

<sup>2</sup><https://data.sdss.org/sas/dr16/eboss/lss/catalogs/DR16/>

<sup>3</sup>In order to distinguish this work from the Ly  $\alpha$  quasar sample, we will denote our sample as ‘LSS quasars’.

presented in Padmanabhan et al. (2008) to Pan-STARRS imaging (Kaiser et al. 2010). The photometry with updated calibrations was released with SDSS DR13 (Albaret et al. 2016) and was demonstrated to have sub-percent level residual calibration errors (Finkbeiner et al. 2016). This DR13 photometric data sample was used to inform the optical selection of eBOSS targets.

The infrared data were obtained using the *Wide Field Infrared Survey Explorer* (WISE; Wright et al. 2010). The WISE satellite observed the entire sky using four infrared channels centred at 3.4  $\mu\text{m}$  (W1), 4.6  $\mu\text{m}$  (W2), 12  $\mu\text{m}$  (W3), and 22  $\mu\text{m}$  (W4). We used the W1 and W2 data to identify eBOSS targets. All targeting is based on the publicly available *unWISE* coadded photometry, which obtained results for SDSS sources via ‘force-matching’ (Lang 2014; Lang, Hogg & Schlegel 2016).<sup>4</sup>

The details of the quasar selection are presented in Myers et al. (2015), where it was demonstrated that SDSS+WISE imaging data can reliably select quasars with  $0.9 < z < 2.2$ . The method combined three essential pieces:

- (i) XDQSOz (Bovy et al. 2012) reporting a greater than 20 per cent chance of an object being a quasar at  $z > 0.9$ ;
- (ii) an extinction corrected flux cut  $g < 22$  or  $r < 22$ ;
- (iii) a mid-IR-optical colour cut, which was proven to be efficient at removing stellar contaminants.

An important aspect of the quasar targets is that many were previously observed in SDSS I/II/III. Such targets are denoted as ‘legacy’; the LSS quasar legacy targets were not re-observed. Legacy targets are not isotropically distributed over the sky and thus must be treated carefully; these details are provided throughout Section 5. Ata et al. (2018) demonstrated that selecting quasars that were subsequently measured to have  $0.8 < z < 2.2$  provided an excellent sample for LSS analyses. Here we will detail how we have built on these results to provide the final eBOSS quasar LSS catalogues. The target density is  $112 \text{ deg}^{-2}$  within the  $6309 \text{ deg}^2$  area planned for eBOSS observation.

The full details of the LRG selection are given in Prakash et al. (2016). The goal was to obtain a sample at redshifts greater than the BOSS CMASS sample. In order to make it distinct, the sample was selected to be fainter in the  $i$ -band than BOSS CMASS galaxies (Reid et al. 2016). Flux cuts were applied in the  $i$ - and  $z$ -bands in order to obtain targets bright enough to achieve a successful redshift. Optical/infrared colour cuts achieved a sample with redshifts mostly greater than  $z = 0.6$ , which is near where the density of the CMASS ceases to produce cosmic variance limited clustering measurements. Bautista et al. (2018) demonstrated the sample to be viable for LSS studies. The target density is  $60 \text{ deg}^{-2}$  within the area planned for eBOSS observation. Here, we provide the details on the final LRG sample and combine it with the high redshift tail of the BOSS galaxy sample in order to provide one larger sample of LRGs with  $z > 0.6$ .

Files containing the LRG and quasar target information applied to the full SDSS imaging were released<sup>5</sup> in DR14 (Abolfathi et al. 2018). They can be matched to the ‘full’ files we describe later (Section 5) using the ‘OBJID\_TARGETING’ column.

## 2.2 ELGs

The eBOSS ELG sample is unique from other eBOSS samples in that it does not use SDSS imaging for its target selection. Instead, ELG targets were selected from the DECam Legacy Survey (DECaLS;

Dey et al. 2019) photometric catalogue. The details of the selection are presented in Raichoor et al. (2017). There, it was demonstrated that within two separate  $\sim 600 \text{ deg}^2$  regions, a  $g$ -band flux cut ( $g < 22.9(22.825)$  in the NGC (SGC) region) and a  $(g - r)/(r - z)$  colour selection were efficient at producing targets over the redshift range  $0.6 < z < 1.1$  with sufficient  $O[II]$  flux to obtain a good redshift. The target catalogue was made public<sup>6</sup> in DR14. Raichoor et al. (2020) present further details on the ELG LSS catalogue construction and its viability for LSS studies and we thus repeat few of them here.

## 2.3 Other targets

Observations of two additional samples of high redshift quasars for Lyman  $\alpha$  forest studies (Blomqvist et al. 2019; Chabanier et al. 2019; de Sainte Agathe et al. 2019) were also conducted during eBOSS. The first consisted of known  $z > 2.1$  quasars where increased signal to noise would lead to improved cosmology constraints. The second program used multi-epoch imaging data from the Palomar Transient Factory (PTF; Law et al. 2009; Rau et al. 2009) to select high-redshift quasar targets at a density of  $20 \text{ deg}^{-2}$  in regions with many epochs of photometry (Palanque-Delabrouille et al. 2016). These provide a random sampling of the foreground distribution of neutral hydrogen and do not require a careful record of the selection function for cosmology studies. The Time Domain Spectroscopic Survey (TDSS; Morganson et al. 2015) and the Spectroscopic Identification of *eROSITA* Sources (SPIDERS; Clerc et al. 2016; Dwelly et al. 2017) programs were also conducted simultaneously with eBOSS observations. The Lyman  $\alpha$ , TDSS, and SPIDERS samples do not directly contribute to the clustering catalogues presented here, but the footprint of these observations is incorporated into the clustering catalogues as will be described below.

## 3 SPECTROSCOPIC OBSERVING

The eBOSS targets were primarily observed using the BOSS double-armed spectrographs (Smee et al. 2013) on the 2.5-metre Sloan Telescope (Gunn et al. 2006) at the Apache Point Observatory in New Mexico, USA. The exception is legacy quasar observations that used the original Sloan spectrograph. Here, we describe the details of the observational strategy and how it impacts our final sample, while defining key terminology. The details of how spectra are turned into redshift estimates are presented in Section 4.

Like BOSS, eBOSS observed 1000 targets at a time through fibres plugged into holes on pre-drilled aluminum plates. Plates were placed at the focal plane of the telescope and the fibres fed directly into the two spectrographs. On each plate, targets cannot be placed within 62 arcsec of each other due to the physical size of the housing of the optical fibre (Dawson et al. 2013). Targets are assigned to ‘collision groups’ via a ‘Friends-of-Friends’ algorithm with a 62 arcsec linking length (Reid et al. 2016). Any instance where a target is not observed because it is in a collision group is recorded as a ‘fibre collision’. A fraction of these collisions can be resolved in regions of overlapping plates. The ‘tiling’ algorithm (Blanton et al. 2003) determines the number of plates and the location of plate centres in celestial coordinates. In BOSS, tiling over a fixed area produced a near-optimal solution of field locations that guaranteed 100 per cent completeness of non-collided targets for the primary clustering samples. In eBOSS, the 100 per cent completeness requirement was relaxed for the LRG sample to increase the fibre

<sup>4</sup>These data have since been improved as described in Meisner et al. (2019).

<sup>5</sup><https://data.sdss.org/sas/dr14/eboss/target/ebosstarget/v0005/>

<sup>6</sup><https://data.sdss.org/sas/dr14/eboss/target/elt/decals/>



efficiency and total survey area. In both BOSS and eBOSS, each fixed area that was tiled in a single software run is referred to as a ‘chunk’. The eBOSS LRG sample had a completeness of non-collided targets exceeding 95 per cent in every relevant chunk of the survey. Areas covered by a unique set of plates are ‘sectors’. Completeness statistics are determined on a per-sector basis.

LRG and LSS quasar targets were observed on the same plates, along with targets from the TDSS and SPIDERS programs. TDSS and SPIDERS were each allocated an average surface density of targets of approximately  $10 \text{ deg}^{-2}$ . These plates also contained fibres allocated to the two Lyman  $\alpha$  quasar target samples. It is possible for any of these samples to overlap in targeting with any other. Considering one particular sample, e.g. LSS quasars, the fact that it passed another sample’s criteria is generally ignored; i.e. it is simply treated as any other LSS quasar when constructing the LSS quasar catalogues. When the target selection criteria are distinct, we must consider the effect of fibre collisions where LRG or quasar targets could not be allocated fibres because of these additional targets.

Fibre collisions between different target categories were resolved based on the following priority: SPIDERS, TDSS, re-observation of known quasars, LSS quasars, and variability-selected quasars, with LRGs last. The fibre collision areas occupied by higher priority targets are treated in a ‘veto mask’ that removes area from the window function of the desired clustering sample. Thus, these priorities result in an LRG sample that covers substantially less total area than the quasar sample, despite being observed at the same time across the same large-scale footprint. See Section 5.2 for more details.

A significant portion of the LRG and quasar targets was observed in the Sloan Extended QUasar, ELG and LRG Survey (SEQUELS) that was designed as a pilot survey for eBOSS (see Dawson et al. 2016 for details). We treat SEQUELS targets that pass the eBOSS target selection the same as eBOSS observations, as the selection is a simple super-set of the ultimate selection. A list of the chunk numbers is given in Table 1. The SEQUELS targets are covered by chunks boss214 and boss217.

ELGs were observed separately from the quasars and LRGs in chunks eboss21, eboss22, eboss23, and eboss25. Some TDSS targets shared their plates and were given equal priority. While they were observed in separate chunks, the ELG footprint overlaps with the LRG and quasar footprints and thus allows cross-correlation studies (e.g. Alam et al. 2019; Wang et al. 2020; Zhao et al. 2020b).

Although each chunk was tiled independently, there are some small regions of overlap in area. In other words, along some chunk edges, targets were assigned to more than one chunk. To account for this duplication, we removed the overlap area from the greater numbered chunk. Doing so provides a unique set of targets over the full eBOSS footprint. However, we always take the highest quality spectrum for duplicate tilings of the same target. For example, if a target received a fibre in both chunks eboss20 and eboss26 and the better spectrum was observed in eboss26, the redshift from the eboss26 observation is assigned. However, if the target in eboss20 was not assigned a fibre (e.g. due to a collision), but gets observed in eboss26, the eboss26 observation is not used in the clustering catalogues. This avoids biasing the selection probabilities within these regions. Such cases are fairly rare and are treated as if no fibre was placed on the target. The geometry file described in Section 5 cuts between chunks at the boundary of the highest-numbered chunk, corresponding to this selection.

In many cases, eBOSS chunks that were tiled did not have all of their plates observed. This is the primary source of incompleteness in the eBOSS catalogues. We denote the following classifications that lead to incompleteness for an eBOSS target within a tiled chunk:

**Table 1.** The eBOSS chunks, the number of tiles in the chunk, and the number of tiles with good observations in the chunk. (See the text for definitions of ‘chunk’ and ‘tile’; see also table 1 of Reid et al. 2016.) boss214 and boss217 were SEQUELS chunks, while chunks 21, 22, 23, and 25 were ELG plates. All others were standard LRG+QSO plates. SEQUELS tiled further chunks, which were unobserved at the end of the survey: the area in these chunks was recovered by eBOSS chunks, but without any overlap in the tiles.

chunk	# of tiles	# of good tiles
boss214	148	88
boss217	74	29
eboss1	199	195
eboss16	128	127
eboss2	98	81
eboss20	42	42
eboss21	46	46
eboss22	121	121
eboss23	87	85
eboss24	81	51
eboss25	51	51
eboss26	171	76
eboss27	94	37
eboss3	204	180
eboss4	80	80
eboss5	70	70
eboss9	34	34

(i) ‘close-pair’: No fibre was placed on the target due to a fibre collision (unresolved with overlapping plates) with a target of the same class.

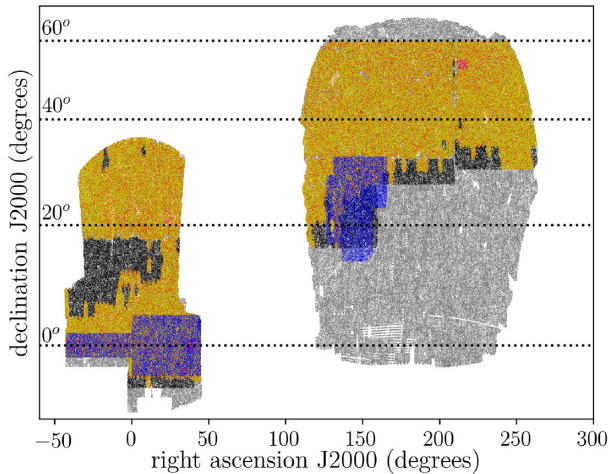
(ii) ‘missed’: No fibre was placed on the target, not due to it being a close-pair. Observations will be missed primarily due to missing plates in overlap regions and can also occur when more than 1000 fibres would have been required to observe all targets in a given region.

(iii) ‘wrong-chunk’: A fibre was placed on the target only in the greater-numbered overlapping chunk.

Fig. 1 displays the sky positions of eBOSS targets that were tiled. The black points are LRG and quasar targets that were not observed. The coloured points display eBOSS targets observed with plates determined to be ‘good’. The red points are LRG targets that were observed. They are mostly overlapped by the yellow points, which show the quasars that were observed. One can observe an area at RA, Dec  $\sim 225, 55$  where there are no quasars. In this area, SDSS had previously obtained spectra for all of the quasar targets (incorporated into the special Reverberation Mapping program; Shen et al. 2015) and thus eBOSS did not re-observe them. The blue points display 20 per cent of the ELGs that were observed. This subsampling allows one to see the overlap with the LRG and quasar samples. The overlap of the ELG sample is complete in the South Galactic Cap (SGC; the filled area in the figure with Right Ascension  $< 60$ ). In the North GC (NGC), the ELG data fully overlaps with the BOSS CMASS data, but the eBOSS LRG and quasar footprint only covers approximate half of the NGC ELG footprint.

#### 4 DETERMINING REDSHIFTS

The IDLSPEC2D spectral reduction pipeline reduces every eBOSS spectrum from a series of 2D images that span multiple exposures, to a single, wavelength-calibrated, 1D spectrum. The spectra that were



**Figure 1.** The footprint of eBOSS targets. The black points show LRG and quasar targets that were tiled but did not obtain spectroscopic observations (see the text). The yellow points show quasars that were observed. They almost entirely overlap the red points, which show LRGs that were observed. The blue points show 20 per cent of the ELGs that were observed. The grey points are BOSS CMASS galaxies from their LSS catalogues. The CMASS data has their veto masks applied, while no such masks are applied for the eBOSS data in this plot. The eBOSS LRG and quasar footprints with veto masks applied are shown in Section 5.4 and the equivalent ELG footprint is shown in Raichoor et al. (2020).

used to generate the catalogues presented in this paper were processed with version v5.13.0 of the data reduction pipeline. This is the final version of the IDLSPEC2D software that will be used to process clustering data obtained with the SDSS telescope. A summary of the improvements to this software package over the course of eBOSS can be found in the studies that first incorporated those improvements (Hutchinson et al. 2016; Jensen et al. 2016; Bautista et al. 2017) and in the DR16 paper (Ahumada et al. 2020).

As in SDSS and BOSS, every spectrum is then assigned a classification of star, galaxy, or quasar, a redshift, and a quality flag that indicates the robustness of the redshift estimate. The redshift catalogues associated with DR16 exactly follow the procedures described in Albareti et al. (2016) and Bolton et al. (2012). However, different philosophies for redshift estimates and spectral classification were designed specifically for the eBOSS LSS catalogues. A new redshift estimate pipeline for galaxies was motivated by the challenges faced with the low signal-to-noise galaxy spectra. A new scheme that supplemented automated classifications with visual inspections was developed to characterize the very large number of quasar spectra obtained in eBOSS. We describe the new algorithms customized to LRG spectra in Section 4.1 and briefly summarize the procedures for ELG and quasar spectra in Section 4.2 (these are described in greater detail in Lyke 2020; Raichoor et al. 2020).

The relative success of classification is divided into three cases: good redshift, redshift failure, and no chance of good redshift (‘bad fibre’). For all three LSS tracers, the bad fibres are determined based on the ZWARNING flag from the eBOSS pipeline. Observations with bits 1 (‘LITTLE COVERAGE’), 7 (‘UNPLUGGED’), 8 (‘BAD TARGET’), or 9 (‘NO DATA’) had no chance of obtaining a good redshift and are classified as bad fibres. As the cases of bad fibres are uncorrelated with the target properties, they are treated in the same manner as if they did not receive a fibre in the catalogue creation, as described in Section 5.4. The following subsections detail how we

classify between good redshifts and failures for LRGs and quasars. We describe the characterization of the spatial variation of redshift failures and our statistical corrections for them in Section 5.3.

#### 4.1 Redrock redshift estimates for the LRG sample

As discussed in Dawson et al. (2016), the spectra from the BOSS CMASS galaxy sample had sufficient signal to noise to enable very reliable automated redshift classification using the same algorithms as those in the recently released DR16 catalogues. However, early in SDSS-IV, it became clear that these routines are not optimized for the fainter, higher redshift LRG galaxies that comprise the eBOSS LRG sample. When first applied to the eBOSS samples, only about 70 per cent of the spectra were given good redshifts. The high rate of redshift failures motivated the new development in the IDLSPEC2D spectral reduction pipeline for higher quality 1D spectra. More significant improvements to the rate of good redshift estimation were achieved through a new approach to redshift estimation.

The new redshift algorithm, REDROCK,<sup>7</sup> was developed for the Dark Energy Spectroscopic Instrument (DESI; Aghamousa et al. 2016a). The REDROCK team used an improved combination of the Bolton et al. (2012) approach and an archetype (Cool et al. 2013) approach similar to that applied in REDMONSTER (Hutchinson et al. 2016). Methods developed in Zhu (2016)<sup>8</sup> were incorporated in order to provide additional improvements. We describe the approach in more detail throughout the rest of this section.

The general process, which we expand on below, is as follows: Classification and redshift determination are performed via a fit of a linear combination of spectral templates to each spectrum. Fitting is done over a range of redshifts for three different classes of templates that independently characterize stellar, galaxy, and quasar spectral diversity. Unlike the approach used in the BOSS redshift pipeline, no nuisance terms are allowed to soak up flux calibration errors, intrinsic dust extinction, or other sources of spurious signal. The redshift and spectral class that give the lowest value of  $\chi^2$  are considered the best description of the spectrum. A fit is only considered reliable, or good, if it can be differentiated from the second best fit by a sufficiently large difference in the  $\chi^2$ . We denote this parameter as  $\Delta\chi^2$ .

The first improvement over the BOSS fitting routines was the introduction of the instrument resolution to the spectral models. Each model is generated at a significantly higher resolution than offered by the BOSS spectrograph. At each redshift, the model is convolved with the wavelength-dependent estimate of the Gaussian profile that describes the instrument resolution for that spectrum. The inclusion of instrument performance in this step allows better characterization of narrow spectral lines, particularly when there is a strong variation in the resolution as a function of wavelength as often occurs near the detector edges.

The second improvement over the BOSS fitting routines is an introduction of new spectral templates for galaxies and stars.<sup>9</sup> Galaxy spectral templates are derived from a principal component analysis (PCA) decomposition applied to a total of 20 000 theoretical galaxy spectra (Charlie Conroy 2014, private communication) that span stellar age, metallicity, and star formation rate.<sup>10</sup> Emission lines of

<sup>7</sup><https://github.com/desihub/redrock>; tagged version 0.14.0

<sup>8</sup>Parts of this associated code were used: <https://github.com/guangtunbenzh/u/SetCoverPy>.

<sup>9</sup><https://github.com/desihub/redrock-templates>; tagged version 2.6

<sup>10</sup>Specifically, these are broken up by DESI target class to have 10 000 ELGs, 5000 LRGs, and 5000 spectra representing the flux-limited ‘Bright Galaxy Sample’

varying equivalent width were painted on to the theoretical galaxy spectra. The resulting PCA eigen spectra are therefore physically motivated, as opposed to the BOSS eigenspectra that were derived empirically from early data and are thus degraded from noise and occasional spurious signal in the spectra. There are 10 galaxy PCA eigenspectra templates that are used in linear combination to obtain redshifts for the entire eBOSS galaxy sample.

The stellar templates were also derived from a series of theoretical models divided approximately by stellar mass and evolutionary stage. The stellar templates were motivated by laboratory atomic data, molecular data, and model atmospheres (Allende Prieto et al. 2018; Allende-Prieto et al., private communication). A total of 30 000 template stars were used. 10 000 had spectral types A, B, F, G, K, or M. 20 000 white dwarf templates were used, split evenly between types DA and DB. Broad TiO absorption features in red dwarf spectra can masquerade as *G*-band or Balmer breaks in high redshift galaxies. Thus, extra care was taken to increase the diversity of M-type and K-type main-sequence stellar templates. The introduction of these new templates was proven to reduce the rate of false detections around  $z = 0.62$  and  $z = 1.02$ . The CV-type stellar templates and the four quasar eigenspectra produced by Bolton et al. (2012) and used in previous eBOSS analyses were copied into REDROCK. The redshifts for the LSS quasar sample were determined as detailed in the following subsection (not by REDROCK).

In the second element of the REDROCK redshift classification scheme, a subset of the spectral templates described above were used as archetype models<sup>11</sup> to fit the spectra in a manner similar to REDMONSTER. The motivation for this second step was to apply an additional filter on the spectral fitting and exclude non-physical combinations of the eigenspectra that can generate erroneous redshift detections. Archetype fitting was not performed over the full redshift range, but instead was performed only over the range of within  $10\,000\text{ km s}^{-1}$  of the redshift estimate from a maximum of the three best-fitting cases for each class (galaxy, quasar, star) from the first stage of classification. For the redshift ranges where the spectral class was estimated to be a galaxy, 110 archetype galaxy templates were fit in combination with nuisance terms that control the amplitude of the first three Legendre polynomials, meant to fit non-physical flux in the broad-band spectrum. Likewise, 40 stellar archetype spectra were fit to the spectrum for the redshifts where the PCA spectral class was estimated to be stellar, and 64 archetypes were fit for class quasar. The redshift and class that produced the lowest value of  $\chi^2$  was then considered the best description of the spectrum. The results from the archetype fits superseded those from the PCA eigenspectra fitting and are used for the clustering catalogues. The  $\Delta\chi^2$  between the best two archetype fits was recorded and later used to define our redshift failure criteria.

To limit the number of interlopers in our LSS measurements, we established a requirement that limited the number of misclassified, or ‘catastrophic failures’, to be less than 1 per cent. A catastrophic failure for galaxies occurs when an object is confidently assigned a redshift that is in error by more than  $1000\text{ km s}^{-1}$ . The final tuning to discriminate between good redshifts and redshift failures and to assess the resulting rate of catastrophic failures was done empirically using multi-epoch spectra and sky spectra, as described below.

We use a sample of multi-epoch spectra to identify a value of the  $\Delta\chi^2$  that maximizes the number of good redshifts while maintaining sufficient purity in the catalogue. Many of these objects received

more than one observation due to intentional reobservations of a plate while others had multiple fibre assignments in the regions of plate overlap. There were 11 556 pairs of spectra used to perform this test. For each pair of spectra, we determined the difference in the redshift estimates,  $\Delta v$ . The distribution of  $\Delta v$  is shown in the left-hand panel of Fig. 2 while the results as a function of  $\Delta\chi^2$  are presented in the right-hand panel. Using the fit to the distribution, which we have cut to the  $0.6 < z < 1.0$  redshift range used for the clustering catalogues, the mean redshift uncertainty for the LRG sample is  $65.6\text{ km s}^{-1}$  ( $1/\sqrt{2}$  the width of the distribution in Fig. 2). An uncertainty of this scale is small compared to typical peculiar velocities and is thus absorbed into their modelling in the LSS analyses.

We then assessed the rate of catastrophic redshift failures by counting the fraction of pairs that produced redshift estimates differing by more than  $1000\text{ km s}^{-1}$ . For a threshold  $\Delta\chi^2 = 9$  (rejecting 761 pairs), we find that 0.5 per cent of the 10 795 pairs produced a catastrophic redshift failure. Under the assumption that one of the redshift estimates in the pair was correct, the resulting catastrophic failure rate is estimated to be 0.25 per cent.

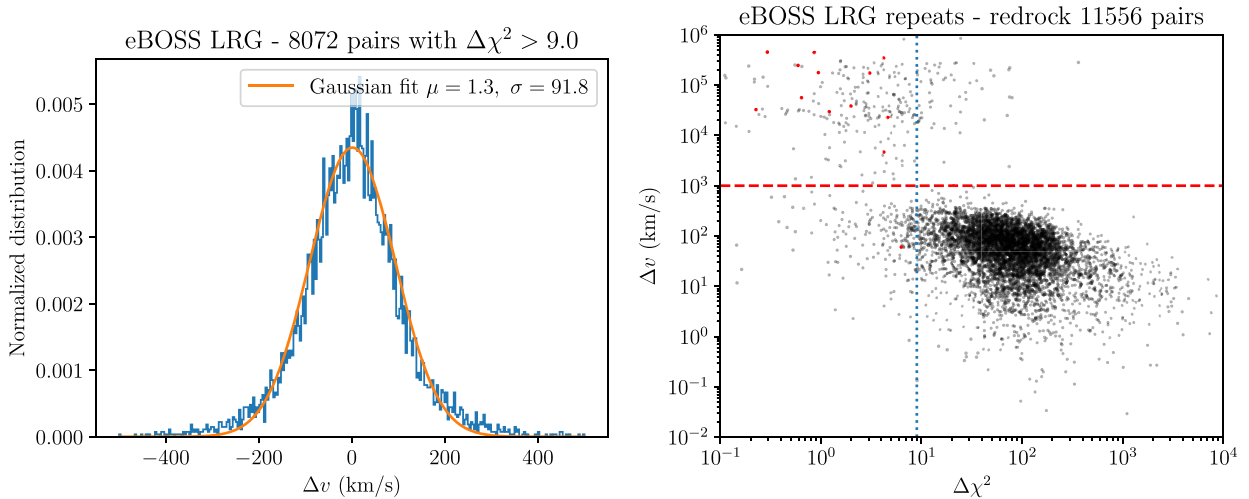
We then applied an additional level of filtering to further reduce the rate of catastrophic failures, which was to require a positive amplitude for the coefficient of the best-fitting archetype spectrum. In cases where the best-fitting redshift was produced by an archetype template with a negative amplitude, we kept that redshift but set a flag indicating that the redshift was not to be trusted. These are counted as redshift failures in the downstream analysis. We applied this condition based on tests of 365 243 sky-subtracted sky spectra. Without the requirement, we found that 10 per cent of these sky spectra were given a confident redshift estimate<sup>12</sup> using the  $\Delta\chi^2 = 9$  threshold, whereas one would expect a negligible fraction of astrophysical spectra in those fibres. With the requirement, this was reduced to 4.4 per cent. While the positive-archetype requirement provides a significant improvement, this behaviour indicates a systematic bias in the algorithm in the limit approaching zero signal. We have not been able to identify the source of this bias. The spectra failing to meet the physicality condition are shown in red in Fig. 2. One notes that these pairs are most likely found at low values of  $\Delta\chi^2$ , as would be expected. There are only 0.04 per cent of LRG spectra in the full eBOSS sample that satisfy the  $\Delta\chi^2 = 9$  condition but fail to meet this threshold on positive archetype coefficients. Given the results on the sky spectra, we can expect a similar percentage of catastrophic failures in our LRG sample due to these false-positive confident redshifts; i.e. this is a negligibly small fraction. The requirement that the first coefficient be positive for the best-fitting archetype spectrum removes spurious detections from non-physical fits to the data, albeit at a very low rate.

After final classifications, the redshift completeness now approaches 98 per cent for the eBOSS LRG sample with a rate of catastrophic failures estimated to be less than 1 per cent. These cases of catastrophic failures appear in the clustering catalogues without correction but are shown to be sufficiently rare as to not bias the cosmological measurements. Stars are a major contaminant, as they make up 9 per cent of the spectral classifications for the LRG sample. An additional one per cent of spectra are classified as quasars and not used in the LSS catalogues. In total, 88 per cent of LRG observations result in a good LRG redshift.

<sup>12</sup>These ‘redshifts’ broadly sample the allowed redshift range with only minor structure that appears to be caused by confusion from sky subtraction artefacts.

<sup>11</sup><https://github.com/desihub/redrock-archetypes>; tagged version 0.1





**Figure 2.** Left-hand panel: The distribution of  $\Delta v$  over 8071 pairs of observations of the same LRG target at  $0.6 < z < 1.0$  and with confident detections ( $\Delta\chi^2 > 9$ ).  $\Delta v$  is the difference in velocity between two redshift measurements of the same object. The solid line shows the best-fitting Gaussian model to the distribution after requiring  $|\Delta v| < 250 \text{ km s}^{-1}$ . The mean and dispersion are shown in the legend. Right-hand panel:  $\Delta v$  as a function of  $\Delta\chi^2$ , for 11 556 pairs where we apply no cut on redshift.  $\Delta\chi^2$  represents the statistical difference between the best and the second best-fitting spectral template to a single spectrum. The lower  $\Delta\chi^2$  in the pair is considered the independent parameter. Pairs in which one spectrum was fit with an archetype spectral template with negative amplitude are presented in red. The horizontal red dashed line shows the limit of  $\Delta v = 1000 \text{ km s}^{-1}$  above which a pair is considered as a catastrophic failure. The dotted vertical line shows the  $\Delta\chi^2 = 9$ , below which results are classified as redshift failures.

#### 4.2 ELG and quasar redshift estimates

We also utilize the REDROCK code to make redshift estimates for the ELG spectroscopic sample. The PCA and the archetype spectral templates are identical to those described above. The requirement for  $\Delta\chi^2$  between models and the restriction on the archetype coefficients are also identical. However, two additional criteria are applied to the ELG program: the median signal to noise per pixel must exceed 0.5 in either the  $i$ -band or  $z$ -band region of the spectrum and the measured continuum or [O II] emission line strength must also pass the a posteriori flags defined and motivated in Comparat et al. (2016) and Raichoor et al. (2017); the criteria using these flags is ( $zQ > 1$  or  $zCont > 2.5$ ). The details of purity and completeness after each of these filters is presented in Raichoor et al. (2020). We are able to obtain secure redshifts for 91 per cent of ELG observations with a catastrophic failure rate of less than 1 per cent.

We use a multistage process to determine the redshift and quality indicator for the quasar sample. This process follows on the philosophy of Pâris et al. (2018) and is fully described in Lyke (2020), which presents the ‘DR16Q’ quasar catalogue. From these results, we used the following criteria to determine redshift failures: if an object was not classified as a quasar by the automated decision-tree described in Lyke (2020)<sup>13</sup> and had an IDLSPEC2D ZWARNING flag set (not associated with the bad fibres described above), it was typed as a redshift failure. If no ZWARNING flag was set, the observation was assigned the classification determined by the IDLSPEC2D pipeline. Additionally, anything with a median signal to noise  $< 0.5$  per pixel across the spectrum was classified as a redshift failure. All redshifts we use in the LSS catalogues were determined using the REDVBLUE<sup>14</sup> principle component analysis (PCA) algorithm described in Lyke (2020) and stored in the Z.PCA

column within DR16Q. Within our redshift range of  $0.8 < z < 2.2$ , we find this redshift performs well both in terms of systematic and statistical uncertainties, as discussed below.<sup>15</sup> Visual inspection information is only used to evaluate the catastrophic failure rate, as described below.

The process results in 95 per cent of quasar target observations having a good redshift with a quasar, stellar, or galaxy classification. Seven per cent of the observations are typed as galaxies and two per cent stars. In total, 86 per cent of the eBOSS quasar observations are classified as having a good quasar redshift.

The statistical uncertainties in quasar redshift estimates are computed empirically using repeat observations. Lyke (2020) find a typical statistical redshift error of  $300 \text{ km s}^{-1}$  without strong redshift dependence. Systematic errors in redshift estimates are somewhat more difficult to assess, as the emission lines that inform the fits are subject to internal dynamics and can be shifted with respect to the quasar rest-frame. Lyke (2020) study this by using results of repeat observations from the Reverberation Mapping program (Shen et al. 2015) and find no evidence of a systematic uncertainty in the PCA redshifts with the range  $0.8 < z < 2.2$ ; see their fig. 3.

Catastrophic failures are characterized via the 10 000 random visual inspections described in Lyke (2020). From this set of 10 000, we select the LSS quasar targets that were classified as quasars and had eBOSS (not legacy) redshifts  $0.8 < z < 2.2$ . This sample provides a base set of 5449 objects that we include as good quasar redshifts in our clustering catalogues. Of these, the visual inspection found 1.2 per cent (63) were not quasars.<sup>16</sup> An additional 0.8 per cent (45) were determined to have redshifts with  $\Delta v > 3000 \text{ km s}^{-1}$  relative to

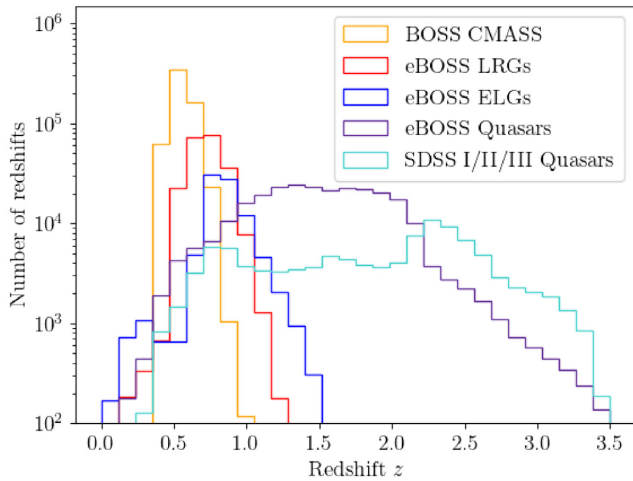
<sup>13</sup>This classification is stored in the column named ‘MY.CLASS\_PQN’ in the ‘full’ quasar catalogue files.

<sup>14</sup><https://github.com/londumas/redvsblue>

<sup>15</sup>For the Ly  $\alpha$  forest studies presented in du Mas des Bourboux et al. (2020) Z.LYAWG is used instead, where Ly  $\alpha$  emission is masked. This is less of a concern in our redshift range.

<sup>16</sup>No accurate new classification or redshift estimate was attempted but any resulting redshift would have been unlikely to be close to the original ‘quasar’ redshift.





**Figure 3.** Histograms of the redshifts of samples used in eBOSS LSS analyses. The quasars are selected to pass our LSS sample target selection, as explained in the text, but many were already observed by previous generations of SDSS. The SDSS-III BOSS CMASS sample is included, as we combine this sample with eBOSS LRGs to produce one larger sample.

the REDVSBUE redshift. These combine for an estimated 2 per cent catastrophic failure rate on LSS quasar targets observed by eBOSS. All legacy redshifts had been visually inspected prior to eBOSS and determined to be good quasar redshifts. Legacy quasars make up 18 per cent of the quasar redshifts used in the LSS catalogues with  $0.8 < z < 2.2$ . We thus estimate the total catastrophic failure rate to be 1.6 per cent for the LSS quasar sample (as the total fraction is  $0.02 \times 0.82$ ).

This statistical characterization of the distribution of redshift on uncertainties in our LSS quasar catalogue is used in Smith et al. (2020) to create simulations that are consistent with these results. Thus, Hou et al. (2020), Neveux et al. (2020) are able to determine the sensitivity of their BAO and RSD measurements to such redshift uncertainties and catastrophic failures and incorporate the results into their systematic error budgets.

### 4.3 Redshift distributions

Fig. 3 displays redshifts from the samples used to create eBOSS LSS catalogues. The SDSS I/II/III quasars were selected as eBOSS LSS quasar targets. These already had secure redshifts determined by visual inspection. Thus, for the LSS quasar sample, we use their previously observed spectra and redshift estimates rather than re-observe them. See Section 2 for more details. These legacy quasars span the whole redshift range and comprise approximately one quarter of the total quasar redshifts. The BOSS galaxies were not targeted by eBOSS, but we will use BOSS CMASS galaxies at  $z > 0.6$  in order to create one combined sample of luminous galaxies with  $z > 0.6$ .

## 5 CATALOGUE CREATION

In this section, we detail the catalogue creation steps for the LRG and quasar samples. The steps are similar for the ELGs, but those catalogues are described in Raichoor et al. (2020). The order for catalogue creation is:

- (i) create randoms at constant surface density within the tiled footprint;

- (ii) match between targets and spectroscopic observations;
- (iii) apply veto masks;
- (iv) resolve fibre collisions and determine completeness;
- (v) assign weights to correct for fibre collisions and redshift failures;
- (vi) cut on redshift and completeness;
- (vii) assign weights that correct for systematic trends with foregrounds and imaging meta data;
- (viii) assign redshift related information to random catalogues.

Many of these steps apply to both the data catalogue and the random catalogue that is used to quantify the window function. The first operation is therefore to create a catalogue with random angular positions at a density of  $5000 \text{ deg}^{-2}$  within the geometry of the full tiled area, which is more than 40 times greater than the target density of the quasar sample (and more than 70 times greater than the LRG target density). This area is a collation of the previously described chunks (with overlap removed) and occupies  $6309 \text{ deg}^2$ . A polygon file that can be used with MANGLE (Swanson et al. 2008) named ‘eBOSS\_QSOandLRG\_fullfootprintgeometry\_noveto.ply’ defines this geometry, with a corresponding FITS<sup>17</sup> file that allows a mapping between polygons and sectors.

For each tracer, we release files containing the following types of tables:

- (i) A table with a row for every unique target that was tiled and passes the veto masks; we denote these the ‘full’ files. They contain all of the information on the target’s photometry and spectra (if observed) and relevant IDs. One can use the column ‘OBJID.TARGETING’ to match to ‘objID’ in the publicly available photometry<sup>18</sup> and the columns ‘PLATE’, ‘MJD’, ‘FIBERID’ to match to the publicly available spectra.<sup>19</sup>
- (ii) A table with rows for only the data with good redshifts, with all mask, completeness, and redshift cuts applied. It contains only the columns that are necessary for calculating two-point statistics and matching to the full file. We denote these the ‘clustering’ files.
- (iii) A table of random points approximating the selection function of the clustering file for the data, to be processed in the same way as the data file for the calculation of two-point statistics.

The clustering files are produced separately for each Galactic hemisphere.

### 5.1 Matching targets and spectroscopic observations

The galaxy catalogue creation starts from the target sample. The information for all eBOSS targets within tiled chunks is collated. From this master list of eBOSS targets, the target sample in question is selected. Each target sample is then matched to spectroscopic observations. A first step is to cut the spectroscopic information to unique entries per target. For LRGs, this is done by selecting primary spectroscopic observations from the SDSS data base (SPECPRIMARY = 1). For quasars, we use the DR16Q superset (Lyke 2020) catalogue as the source of redshifts. The primary record (PRIM\_REC = 1) is selected.

For the quasars, we first match the legacy targets with their spectroscopic information. These objects were flagged in the target file as having already been observed and thus were removed from consideration by the tiling algorithm. We match these targets to

<sup>17</sup>[https://fits.gsfc.nasa.gov/fits\\_standard.html](https://fits.gsfc.nasa.gov/fits_standard.html)

<sup>18</sup><https://www.sdss.org/dr16/imaging/>

<sup>19</sup>[https://www.sdss.org/dr16/spectro/spectro\\_access/](https://www.sdss.org/dr16/spectro/spectro_access/)

DR16Q, populate the relevant spectral information (redshift, object type, etc.), and denote them as legacy. We then match the remaining targets based on their internal ID. For the LRGs, we go straight to matching based on internal ID.

After this matching, the following classifications are possible, which are stored as an integer value in the ‘IMATCH’ column of the ‘full’ file:

- (i) a target can remain unobserved (IMATCH=0); denoted  $_{\text{missed}}$ ,
- (ii) have a good eBOSS redshift that matches the targeted type (IMATCH=1; denoted  $_{z, \text{eboss}}$ ),
- (iii) have previously been determined to be a quasar with a good legacy redshift (IMATCH=2; denoted  $_{\text{leg}}$ , relevant only for quasar targets),
- (iv) be a star (IMATCH=4; denoted  $_{\text{star}}$ ),
- (v) be a redshift failure (IMATCH=7; denoted  $_{z\text{fail}}$ , see Section 4),
- (vi) be identified as an object of the wrong target type (e.g. an LRG target is identified to be a quasar; IMATCH=9; denoted  $_{\text{badclass}}$ ),
- (vii) have previously been determined to be a legacy star (IMATCH=13; relevant only for quasar targets),
- (viii) be a bad fibre (IMATCH=14, see Section 4),
- (ix) or was not tiled in its target chunk (IMATCH=15, see Section 3).

Objects with IMATCH=14,15 are treated the same as unobserved objects for calculating all subsequent statistics, i.e. we tabulate any quantity with the subscript  $_{\text{missed}}$  including the IMATCH=14 and 15 objects. Some IMATCH=2 objects will get re-assigned as IMATCH=8, based on completeness considerations, as described in Section 5.4. IMATCH 5 and 6 are not used.

## 5.2 Veto masks

After the matching and type assignment, a series of veto masks are applied to the targets and randoms. These masks and statistics describing what they remove are detailed in Table 2. Chunks covering a unique area of 6309 deg<sup>2</sup> were tiled for observation. Approximately 500 deg<sup>2</sup> of the area is vetoed from the quasar footprint and more than 1000 deg<sup>2</sup> is vetoed from the LRG footprint. Four veto masks are applied to each of the LRG and quasars. The bad field, bright star, and bright object masks are the same as applied to BOSS DR12 (Reid et al. 2016). The centrepost mask removes the area at the centre of the plate where no target can be observed (the centrepost pulls the centre of the plate such that its curvature approximately matches the best-focus surface; see Smee et al. 2013).

The infrared bright star mask was applied to the LRG sample, as it was found that many spurious LRG targets exist around these stars. The size of the region that was masked is based on the *WISE* W1 magnitude, unless the 2MASS *K*-band magnitude was less than 2. Around each source, a circular region of 550 arcsec was removed from consideration if either W1 or *K* was less than 2 magnitudes. For fainter sources up to W1 = 8 we applied

$$r_{\text{IRmask}} = (1397.5 - 569.34W1 + 79.88W1^2 - 3.75W1^3) \text{ arcsec.} \quad (1)$$

Based on early data occupying 800 deg<sup>2</sup>, 85 per cent of LRG targets removed by this mask were not LRGs. Thus, the mask was applied to LRG targets used for tiles eboss9 and greater so that the fibres could be assigned to targets more likely to produce good redshifts. These IR stars were not found to have any impact on quasar targets, beyond what is masked by the regular bright star mask.

The collision priority mask removes the 62 arcsec radius area around where higher priority targets prevent any fibre to be assigned to the given target type. The LRGs had the lowest priority and

the area of collision priority mask applied for them is thus nearly 700 deg<sup>2</sup>. The overlapping plate geometry allows collisions between lower priority LRGs and higher priority targets to be resolved. However, these collisions are not fully resolved and some LRGs remain unobserved in these regions. We thus apply the conservative option of masking 62 arcsec around every higher-priority target and accept losing the 10 439 good redshifts in this mask.

Only TDSS and SPIDERS have greater priority than the quasars. Also, the quasar collisions in regions with overlapping plates are fully resolved. Thus, we only apply the quasar collision priority mask in single tile regions. This mask is only 66 deg<sup>2</sup> for the quasars and only 39 of the more than 7000 quasar targets removed by this mask have good redshifts; the number is greater than 0 only due to the fact that the centre of the veto regions are within the single tile region but can extend out into the area with overlapping tiles.

Note that by applying the veto masks to both galaxies and randoms, we are implicitly assuming that the regions removed are uncorrelated with the cosmological density fluctuations that we want to measure. This may be a slight concern where higher priority targets overlap in redshift with the sample of interest. The main concern is clustering between  $z < 1$  quasars and our LRG sample. We apply no correction for this and expect it to be a minor effect on the LRG clustering given the substantially lower projected number densities of  $z < 1$  quasars compared to  $0.6 < z < 1$  LRGs. This is not an issue for the ELG/LRG multitracers analysis, as these samples were observed on different plates and in different chunks, so the observation of one has no impact on the other.

In general, the veto masks were not applied to the target samples. Thus, many good redshifts were observed within these vetoed regions. For example, for the Bright Star and Bad Field masks, we do not trust that the photometry used to produce the target samples should produce isotropic samples suitable for large-scale structure. These areas thus tend to have proportionally fewer good redshifts. Across all veto masks, for LRGs, nine per cent of the good redshifts are vetoed, to be compared to 17 per cent of the tiled area. For the quasars, we lose 4.5 per cent of the good redshifts while removing 7.1 per cent of the tiled area.

## 5.3 Spectroscopic completeness weights

After the veto masks were applied, ‘close pairs’, denoted  $_{\text{cp}}$ , were assigned. Any object without a spectroscopic observation that shares a collision group with an object that obtained a spectroscopic observation is typed as a close pair and given IMATCH = 3. The distributions of these close pairs and also the redshift failures are not expected to be isotropic and close pairs are expected to be correlated with the density field itself. These sources of spectroscopic incompleteness require special treatment.

For the close pairs, the weights are assigned and equally distributed per collision group. All good observations in a collision group receive a weight that is

$$w_{\text{cp}} = \frac{N_{\text{cp}} + N_{z, \text{eboss}} + N_{\text{badclass}} + N_{\text{star}}}{N_{z, \text{eboss}} + N_{\text{badclass}} + N_{\text{star}}}, \quad (2)$$

where the  $N$  are summed within each of these groups. Such a weighting provides unbiased transverse clustering on large scales in configuration space. However, the radial clustering will be biased, and the issues are more severe in Fourier space (Hahn et al. 2017). Bianchi & Percival (2017) provide an unbiased solution for configuration space and Mohammad et al. (2020) present an application of these weights to eBOSS data. However, in the standard catalogues

**Table 2.** Statistics for the LRG and quasar samples within tiled areas and within the veto masks and completeness cuts we apply to obtain the final sample. Within each region, we list the number of targets, the number of good eBOSS (not legacy) redshifts, and its total area. Many of the veto masks have some overlap, but the statistics are presented individually (thus the total vetoed area is less the sum of the area column). The statistics for the region with  $C_{\text{eBOSS}} \leq 0.5$  are presented after the veto masks have been applied and the statistics for  $C_z \leq 0.5$  are after the  $C_{\text{eBOSS}} \leq 0.5$  cut has been applied.

Region	$N_{\text{LRG tar}}$	$N_{\text{LRG z}}$	$N_{\text{QSO tar}}$	$N_{\text{QSO z, eBOSS}}$	LRG Area (deg <sup>2</sup> )	QSO Area (deg <sup>2</sup> )
Full tiled area	377 633	230 935	703 521	340 386	6309	6309
veto masks:						
LRG collision priority	43 450	10 439	—	—	707	—
QSO collision priority	—	—	7159	39	—	66
Bad field	13 542	7825	20 419	11 448	238	238
Bright star	5910	1899	18 497	4493	131	131
Infrared bright star	6583	849	—	—	72	—
Bright object	1745	788	2993	1212	28	28
Centrepot	36	0	204	0	0.6	0.6
Tiled area, after veto mask	311 848	209 894	655 521	325 226	5223	5858
Completeness cuts:						
$C_{\text{eBOSS}} \leq 0.5$	58 575	2044	116 527	1179	978	1047
$C_z \leq 0.5$	53	20	366	32	1.6	3.4
Tiled area, after veto masks and completeness cuts	253 220	207 830	538 628	324 015	4242	4808

we simply provide  $w_{\text{cp}}$  and each individual analysis describes how the size of any remaining systematic biases how they are treated.

We provide corrections for redshift failures based on the spectrograph signal to noise in the  $i$ -band and the fibre ID. The likelihood of obtaining a good redshift naturally correlates with the signal to noise of the spectrum. The fibre ID correlates with the location of the spectrum on the CCD of the spectrograph, which in turn alters the signal to noise of the spectrum. The fibre ID also correlates with the expected location on the plate, resulting in large-scale signal-to-noise variations across the sky.

We fit for trends between these quantities and the redshift efficiency, as defined below, and use the inverse of the trends as a weight. We define the number of good spectra associated with a particular data subsample<sup>20</sup> as

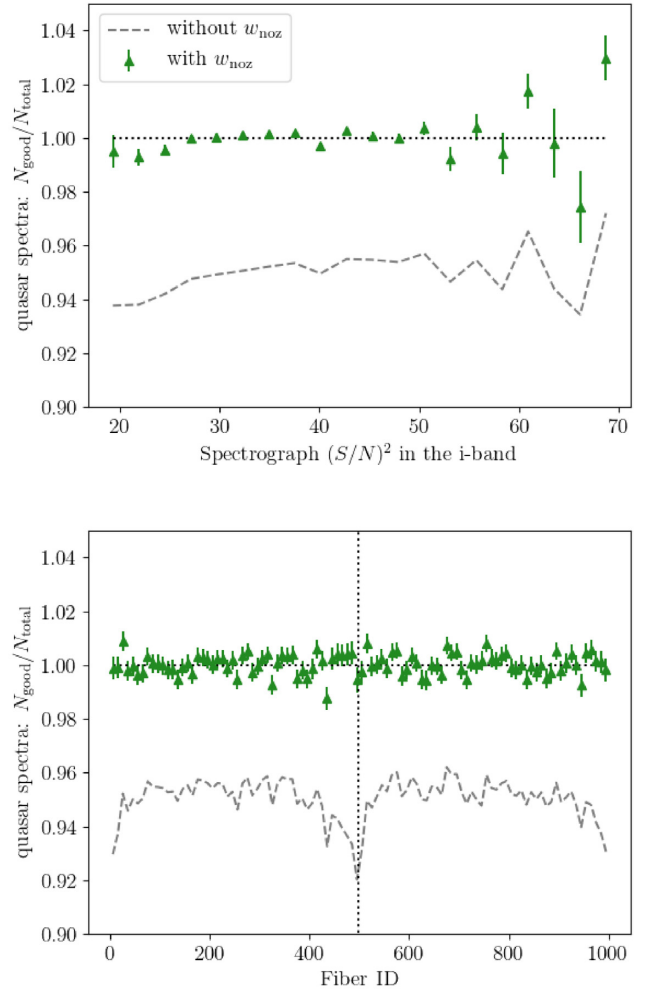
$$N_{\text{goodz}} = N_{\text{z,eBOSS}} + N_{\text{badclass}} + N_{\text{star}}. \quad (3)$$

$N_{\text{alleBOSS}}$  is then  $N_{\text{goodz}} + N_{\text{zfail}}$  and the redshift efficiency for any particular sub-sample is thus

$$f_{\text{good}} = \frac{N_{\text{goodz}}}{N_{\text{alleBOSS}}}. \quad (4)$$

These statistics allow us to characterize the efficiency versus particular aspects of the data (e.g. spectrograph, plate, fiberID) and derive statistical corrections for any trends that we find.

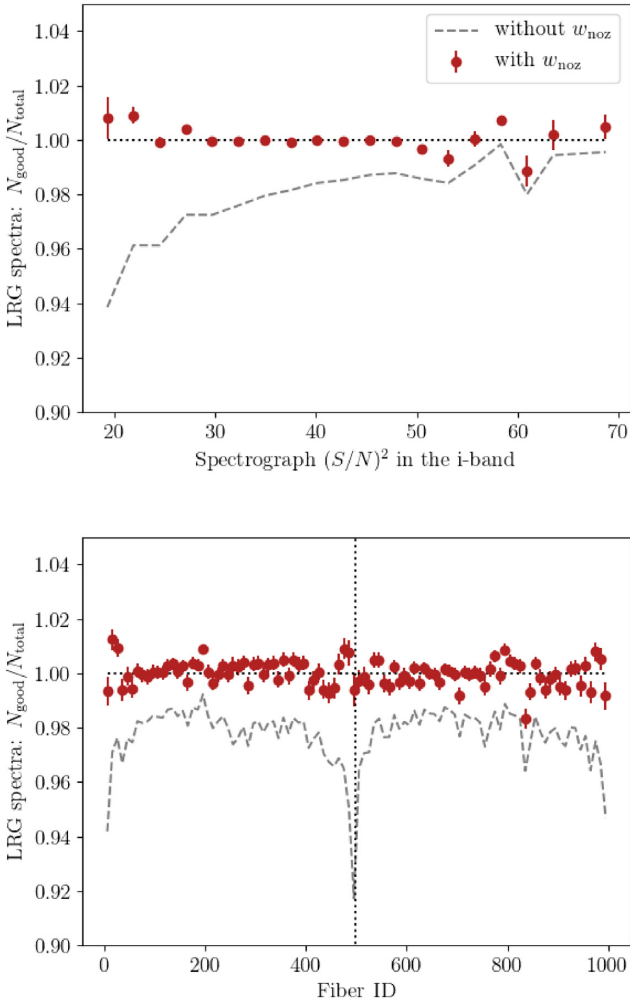
We use the square of the spectrograph signal to noise,  $S_i$ , defined as the square of the median signal to noise of each spectral pixel in the  $i$ -band filter estimated at a photometric magnitude<sup>21</sup>  $i = 20.2$ . This is empirically computed for each spectrograph independently using the combination of all measured spectra from each observation. The trends in redshift efficiency versus  $S_i$  are shown in the bottom panels of Figs 4 and 5, where  $f_{\text{good}}$  is displayed as a function of the spectrograph signal-to-noise ratio in the  $i$ -band,  $S_i$  using dashed grey curves. One can observe that the overall redshift efficiency is higher for the LRGs, but the trend with  $S_i$  is stronger than it is for the quasars. In general, the LRG redshift efficiency is more dependent



**Figure 4.** The fraction of good quasar spectra as function of the square of spectrograph signal to noise in the  $i$ -band ( $S_i$  in the text; top panel) and as a function of the fibre ID (bottom panel). The vertical dotted line at fibre ID 500 denotes the split between spectrographs 1 and 2. The grey dashed lines display the result when not applying the  $w_{\text{noz}}$  weights that we determine based on these quantities, as described in the text.

<sup>20</sup>A subsample can be, e.g. all spectra associated with a spectrograph on a single plate, all of the spectra associated with a given fiberID over all eBOSS observations, etc.

<sup>21</sup>This magnitude is motivated by the typical brightness of an LRG target.



**Figure 5.** The fraction of good LRG spectra as function of the spectrograph signal to noise in the *i*-band (top panel) and as a function of the fibre ID (bottom panel). The grey dashed lines display the result when not applying the  $w_{\text{noz}}$  weights that we determine based on these quantities, as described in the text.

on the signal-to-noise level in a particular spectrum than is the quasar efficiency.

The two classes of tracer have different dependence on exposure depth due to the different spectral features that inform the automated classification. In the case of quasars, most targets have strong emission lines that appear at a considerably higher signal to noise than does the continuum, thus facilitating relatively uniform redshift efficiencies. On the other hand, the primary features used for LRG spectral classification are absorption lines that have a significance determined entirely by the signal to noise of the continuum. Exposure depths were specifically tuned to LRG redshift efficiencies, so variations in  $S_i$  appear as variations in sensitivity to absorption line features and variations in redshift efficiency.

We wish to (statistically) remove these trends from the data so that there is no spurious clustering signal in our catalogues that is associated with plate-to-plate variations in exposure depth. We apply the following steps, which mimic the modelling applied to the DR14 LRG sample in Bautista et al. (2018). We find a linear relationship between  $f_i = N_{\text{good},i}/N_{\text{zfail},i}$  and  $S_i$ , which is determined per spectrograph per plate. One can observe that  $f_{\text{good}} = 1 - 1/(1 + f_i)$ . We perform the fits for each sample in each hemisphere. The fit

is translated to a model for  $f_{\text{good}}$ . Thus

$$f_{i,\text{mod}} = a_S + b_S S_i \quad (5)$$

and

$$f_{\text{good},\text{mod}} = 1 - 1/(1 + a_S + b_S S_i). \quad (6)$$

The inverse of this is used as a weight,  $w_{\text{noz},S}$  to be applied to every good eBOSS observation.

The fibre ID,  $r_{\text{ID}}$ , correlates directly with the location on the CCD where the spectrum is readout. The bottom panels of Figs 4 and 5 use grey curves to display  $f_{\text{good}}$  as a function of  $r_{\text{ID}}$ . Similar trends are observed in both samples, with the dependence again being stronger for the LRGs. The clearest trend is a decrease in redshift efficiency near  $r_{\text{ID}} = 1, 500, 1000$ . These  $r_{\text{ID}}$  correspond to the edges of the CCDs, which are 1, 500 for spectrograph 1 and 501, 1000 for spectrograph 2. A good spectrum is more difficult to obtain near the edges of CCDs, as the optical quality decreases with increasing separation from the centre of the CCD. The  $r_{\text{ID}}$  also correspond indirectly with the location on the focal plane, with low and high number occurring closer to the edge of the plates. Some patterns are also observed near  $r_{\text{ID}} 250$  and  $750$ , where there are also small decreases in the redshift efficiency. These  $r_{\text{ID}}$  correspond to amplifier locations on the CCDs.

We assume a smooth model for the observed trend in redshift efficiencies with  $r_{\text{ID}}$ , which is general enough to capture the trends described above. For each sample, hemisphere, and  $r_{\text{ID}}$  range 1–250, 251–500, 501–750, 750–1000 we fit a relationship

$$f_{\text{good},\text{mod}} = A_r - B_r |r_{\text{ID}} - C_r|^{D_r}. \quad (7)$$

The inverse of this fit is used as a weight,  $w_{\text{noz},r}$ , to be applied to each good eBOSS observation.

The corrections for the expected redshift failure rate per spectrograph,  $w_{\text{noz},S}$ , and per fibre,  $w_{\text{noz},r}$  yield a combined weight:

$$w_{\text{noz}} = w_{\text{noz},S} w_{\text{noz},r}. \quad (8)$$

The weights are only applied to good eBOSS observations, i.e. objects that are not legacy and either have a good redshift or a securely determined alternative type (this is defined by equation 3). We then normalize the weights such that their sum is equal to  $1/f_{\text{good}}$  for the full sample within a hemisphere. Thus, the points with error bars in Figs 4 and 5, which display the results after applying the  $w_{\text{noz},S}$ , fluctuate around  $f_{\text{good}} = 1$ .

Figs 4 and 5 show clear improvement in the trends with  $S_i$  and  $r_{\text{ID}}$ , with some residual scatter. The statistics should follow a binomial distribution, given that each observation can result in a success or failure. The uncertainty in each bin is thus  $\sigma_{\text{fbin}} = [N_{\text{alleBOSS}, \text{bin}} f_{\text{zfail}, \text{bin}} (1 - f_{\text{zfail}, \text{bin}})]^{1/2} / N_{\text{alleBOSS}, \text{bin}}$ . These error-bars are likely underestimated. Random scatter exists in the signal to noise expected in each particular fibre, e.g. due to the fact that each will not have identical throughput. We expect these kinds of variation to be random with respect to position on the focal plane (and thus sky) and we do not attempt to model them. We use these uncertainty estimates to obtain  $\chi^2$  values for the null test that  $f_{\text{good}}$  is constant with  $S_i$  or  $r_{\text{ID}}$ , but given the un-modelled sources of uncertainty and the fact that we do not attempt to account for any covariance between measurement bins, we do not expect  $\chi^2/\text{dof} = 1$ . The  $\chi^2$  numbers are more useful in quantifying the degree of improvement.

We use 20 bins to present the  $S_i$  results. We find that the  $\chi^2$  for the null test with  $S_i$  for quasars decreases from 122 to 60 when the  $w_{\text{noz}}$  weights are applied. As one would expect, the improvement is more dramatic for the LRGs, where the  $\chi^2$  decreases from 744 to 60. The scatter in the residuals, especially at high  $S_i$ , suggests



that the uncertainties are underestimated (as opposed to the high  $\chi^2$  suggesting we have fit the wrong model).

We use 100 measurement bins to present the  $r_{\text{ID}}$  results. For the quasars, the  $\chi^2$  compared to the null expectation improves from 404 to 114. For the LRGs, the  $\chi^2$  improves from 761 to 224. There is a particularly strong outlier at  $r_{\text{ID}} \sim 850$ . Some component of this high  $\chi^2$  is likely due to underestimation of the uncertainty. For both LRGs and quasars, the trends with  $r_{\text{ID}}$  are stronger than those with the physical focal plane positions and the correction for  $r_{\text{ID}}$  removes the trend observed in the  $X$  focal plane position.

When applying both the close pair and redshift failure weights, we obtain an estimate of the number density as a function of redshift that we *would* have achieved from a complete spectroscopic survey that was able to extract good redshifts/reject bad targets with 100 per cent efficiency. The total spectroscopic completeness weight is given by

$$w_{\text{spec}} = w_{\text{cp}} w_{\text{noz}}. \quad (9)$$

Note that this differs from previous BOSS and eBOSS analyses<sup>22</sup>, which defined the weights such that  $w_{\text{spec, old}} = w_{\text{cp, old}} + w_{\text{noz, old}} - 1$  (Reid et al. 2016).

## 5.4 Completeness

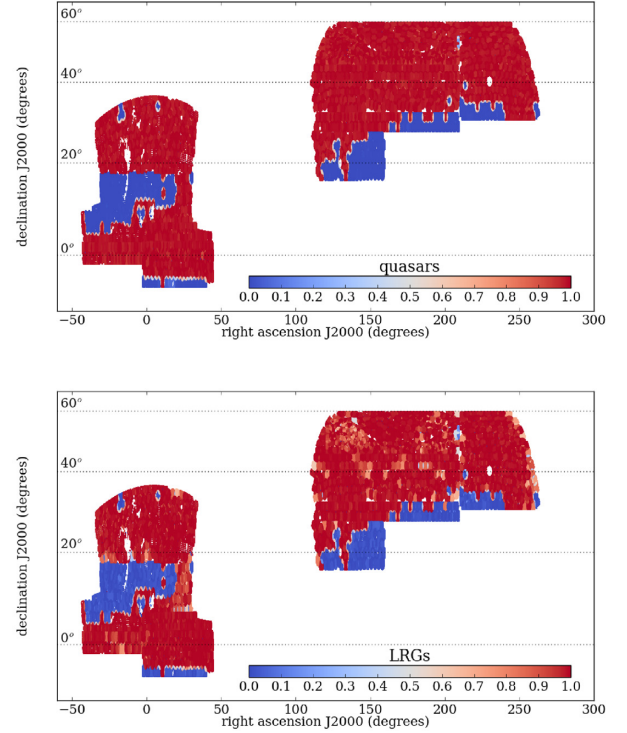
We determine the completeness per sector (each area covered by a unique set of plates) in the same manner as previous BOSS and eBOSS studies:

$$C_{\text{eBOSS}} = \frac{N_{\text{z,eBOSS}} + N_{\text{cp}} + N_{\text{badclass}} + N_{\text{star}} + N_{\text{zfail}}}{N_{\text{z,eBOSS}} + N_{\text{cp}} + N_{\text{badclass}} + N_{\text{star}} + N_{\text{zfail}} + N_{\text{missed}}}. \quad (10)$$

This is basically everything that had a chance of providing a good spectrum, plus close pairs divided by the same plus the remaining number of targets in the sector that were not legacy. This provides us with an angular completeness that is not tied to the instrumental performance (like the redshift failure rate) or the small-scale clustering of the sample in question (like the close-pair weights are). The fluctuations in  $C_{\text{eBOSS}}$  can thus be treated in the random catalogues. Unlike previous BOSS and eBOSS analyses, we do not subsample the random catalogue based on this completeness. Rather, it is used as a weight for all relevant calculations. This has the same effect, with slightly better noise properties. The primary advantage is that one can ignore  $C_{\text{eBOSS}}$  and still calculate any angular statistics on the sample, without any regard to the spectroscopic completeness.

After determining  $C_{\text{eBOSS}}$  for eBOSS quasars, we use it to subsample the legacy observations. This only applies to the quasar sample. The legacy observations represent all quasar targets that had existing SDSS spectra (from any of SDSS I, II, or III). They were removed from the target list sent to the tiling algorithm. Thus, to be in the legacy sample an object must have already been observed and our legacy sample is by definition complete. However, we are weighting the randoms by  $C_{\text{eBOSS}}$  and these randoms are meant to be compared to the full eBOSS quasar sample (including legacy). For this to work, we must artificially impose  $C_{\text{eBOSS}}$  on the legacy sample. To accomplish this, we apply the same choice as previous BOSS and eBOSS analyses: we discard a fraction  $1 - C_{\text{eBOSS}}$  of legacy observations from every sector. In this way, if we now include the remaining legacy observations in  $N_{\text{z,eBOSS}}$  and the discarded ones in

<sup>22</sup>These weights are included in the catalogues as ‘WEIGHT\_CP’ and ‘WEIGHT\_NOZ’.



**Figure 6.** The footprint of eBOSS quasars (top) and LRGs (bottom). The colourmap denotes the completeness. The dark blue areas were not observed by eBOSS (though the areas were tiled).

$N_{\text{missed}}$ , we will recover the same  $C_{\text{eBOSS}}$  as originally determined (and as imparted into the randoms). The discarded objects are assigned IMATCH=8 and included in the ‘full’ catalogues, but are otherwise ignored in the subsequent analyses. This choice allows us to avoid having to include the spatial distribution of legacy observations in the sample mask. From this point onwards, legacy and eBOSS redshifts are treated in exactly the same way.

The completeness of the quasar sectors is shown in the top panel of Fig. 6. One can see that the areas that had any plates observed are highly complete, but there were large areas that were tiled but not observed. Cutting to sectors that have  $C_{\text{eBOSS}} > 0.5$ , the completeness of the quasar sample is 97.7 per cent. The statistics for the quasar sample are presented in Table 3. One can observe that legacy redshifts make up more than one quarter of the total sample, with a higher percentage in the NGC than in the SGC (30 per cent compared to 22 per cent calculated as a fraction of  $N_{\text{eff}}$ ).

The completeness of the LRG sectors is shown in the bottom panel of Fig. 6. One can see that it is nearly the same as that of the quasars, but has lower completeness. While the difference appears significant, the mean completeness when cutting to sectors that have  $C_{\text{eBOSS}} > 0.5$  is only 1 per cent less (96.7 per cent) than that of the quasar sample. The 480 deg<sup>2</sup> decrease in weighted area compared to the quasar sample is not apparent, since this is due almost entirely to the collision priority mask, which removes small holes of radius 62 arcsec. The statistics for the LRG sample are presented in Table 4.

Fig. 7 displays the area in the eBOSS footprint greater than the completeness,  $C_{\text{eBOSS}}$ , shown on the  $x$ -axis. The amount of area with  $0.1 < C_{\text{eBOSS}} < 0.7$  is only 135 deg<sup>2</sup> for the quasars and 144 deg<sup>2</sup> for the LRGs. The threshold applied to the eBOSS clustering catalogues is 0.5, which is shown with the dotted line. This matches the cuts applied to DR14 analyses. Just over 3000 total LRG and quasar

**Table 3.** Basic properties of the quasar LSS catalogues, after veto masks. The quantities are first summed over the full tiled area, with no redshift or completeness cuts. The quantity  $N_{z,\text{tot}}$  is the sum of the legacy and eBOSS redshifts.  $N_{\text{eff}}$  is the effective total number of quasars, after correcting for redshift failures and fibre-collisions, so it is the sum of all good eBOSS and legacy quasars, weighted by  $w_{\text{cp}}w_{\text{noz}}$ . These numbers are reported again after the completeness and then the redshift cuts that are both applied to produce the clustering catalogues. Unweighted area is the sum of the area of all sectors with  $C_{\text{eBOSS}} > 0.5$ ; weighted area multiplies this area by the completeness in each sector and weighted area post-veto multiplies this area by the total fraction of vetoed area. All other quantities are defined in the text.

	SGC	NGC	Total
$N_{\text{eff}}$	177 161	303 298	480 459
$N_{z,\text{tot}}$	165 930	288 522	454 452
$N_{z,\text{eBOSS}}$	126 333	198 893	325 226
$N_{z,\text{legacy}}$	39 597	89 629	129 346
$N_{z,\text{fail}}$	8162	10 616	18 778
$N_{\text{cp}}$	4832	6878	11 710
$N_{\text{gal}}$	9386	18 655	28 041
$N_{\text{star}}$	3758	3327	7085
$N_{\text{star, leg}}$	3830	6669	10 499
after $C_{\text{eBOSS}} > 0.5$ , $C_z > 0.5$ cuts:			
$N_{\text{eff}}$	176 080	302 306	478 386
$N_{z,\text{tot}}$	164 929	287 602	452 531
$N_{\text{eff}}, 0.8 < z < 2.2$	135 244	231 183	366 427
$N_{z,\text{tot}}, 0.8 < z < 2.2$	125 499	218 209	343 708
Area (deg <sup>2</sup> )	1884	2924	4808
Weighted area (deg <sup>2</sup> )	1839	2860	4699

**Table 4.** Basic properties of the LRG LSS catalogues, after veto masks. Quantities are the same as those defined in Table 3 (with  $N_{\text{QSO}}$  the equivalent of  $N_{\text{gal}}$ ).

	SGC	NGC	Total
$N_{\text{eff}}$	87 607	134 695	222 302
$N_{z,\text{tot}}$	82 607	127 287	209 894
$N_{z,\text{fail}}$	2205	3019	5224
$N_{\text{cp}}$	3436	4950	8386
$N_{\text{QSO}}$	1254	1635	2889
$N_{\text{star}}$	10 749	10 017	20 766
after $C_{\text{eBOSS}} > 0.5$ , $C_z > 0.5$ cuts:			
$N_{\text{eff}}$	86 511	133 540	220 051
$N_{z,\text{tot}}$	81 600	126 230	207 830
$N_{\text{eff}}, 0.6 < z < 1.0$	71 427	113 868	185 295
$N_{z,\text{tot}}, 0.6 < z < 1.0$	67 316	107 500	174 816
Area post-veto (deg <sup>2</sup> )	1676	2566	4242
Weighted area post-veto (deg <sup>2</sup> )	1627	2476	4103

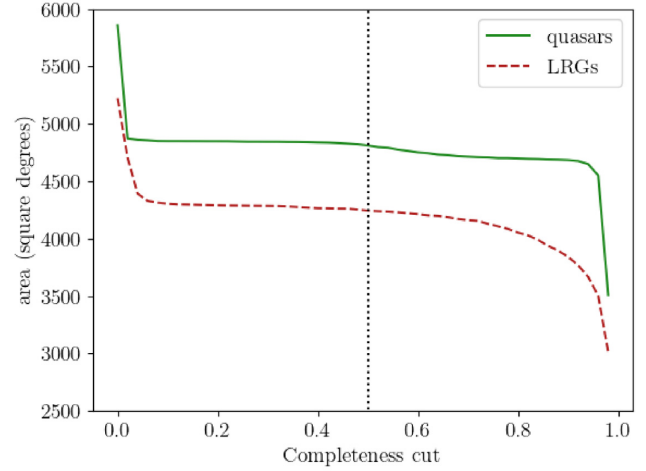
redshifts are removed by this cut, i.e. we lose less than 1 per cent of our eBOSS observations due this completeness cut.

We further track the redshift success rate per sector,  $C_z$ . This is given by

$$C_z = \frac{N_{z,\text{eBOSS}} + N_{\text{badclass}} + N_{\text{star}}}{N_{z,\text{eBOSS}} + N_{\text{cp}} + N_{\text{badclass}} + N_{\text{star}} + N_{z,\text{fail}}}. \quad (11)$$

We will apply a cut  $C_z > 0.5$  to each sample. As shown in Table 2, this cut removes only an additional 3.4 deg<sup>2</sup> for the quasar sample and 1.6 deg<sup>2</sup> for the LRG sample. The change in footprint area as a function of this cut is quite small, as in the range  $0 < C_z < 0.8$  it changes by only 7.5 deg<sup>2</sup> for quasars and by 17.1 deg<sup>2</sup> for the LRGs.

Statistics for the quasars and LRGs after applying the completeness cuts are given in the bottom rows of Tables 3 and 4. At this point, we also determine the  $n(z)$  in the SGC and NGC for each sample. This



**Figure 7.** The area in the eBOSS footprint as a function of the completeness threshold,  $C_{\text{eBOSS}}$ , for quasars and LRGs. The dotted line is at completeness of 0.5, which is the threshold applied to our clustering catalogues.

is shown for the LRGs and quasars in Fig. 8. One can observe that the  $n(z)$  for both samples are significantly different between the NGC and SGC. For the quasars, the difference is primarily a 10 per cent lower density in the SGC that is nearly constant with redshift. This is due to the difference in the mean depth between the two regions. The variations in density versus depth are explored in Section 5.5. For the LRGs, the shapes of the  $n(z)$  are not consistent. The specific differences between the NGC and SGC are not an issue for our samples, as the selection functions are estimated separately for the NGC and SGC. However, systematic variation in the shape of the  $n(z)$  within either region is a systematic concern that we do not treat in our catalogue construction<sup>23</sup> and was not found to be important in the LRG analysis (Bautista et al. 2020). Strong variations in the shape of the  $n(z)$  were found for the eBOSS ELG sample (Raichoor et al. 2020) and found to be important to treat in the RSD analysis (de Mattia et al. 2020; Tamone et al. 2020).

The  $n(z)$  information is used to determine the Feldman, Kaiser & Peacock (1994) weights for the sample

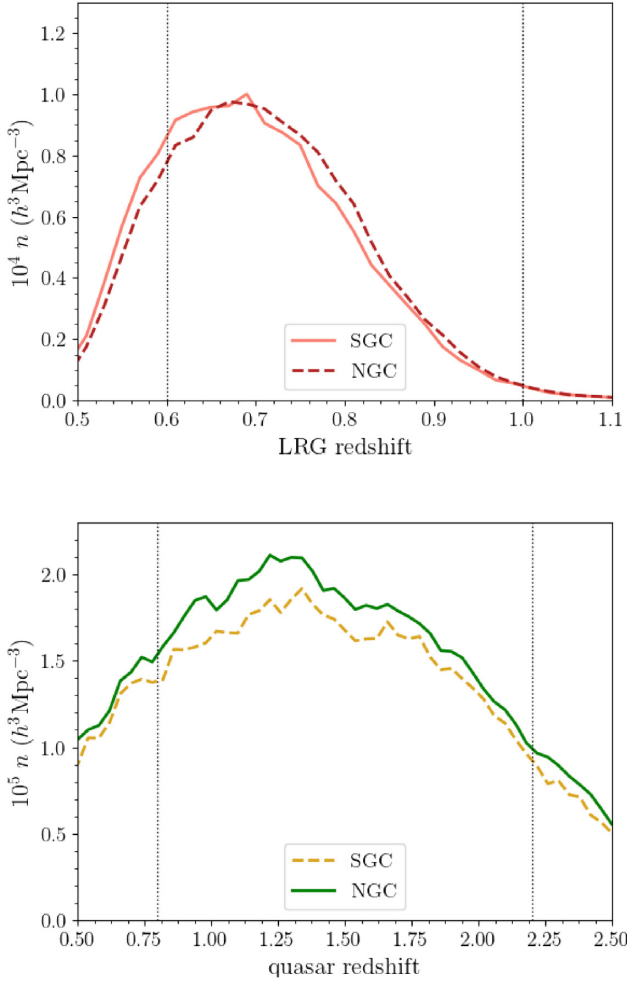
$$w_{\text{FKP}} = 1/[1 + n(z)P_0]. \quad (12)$$

We use  $P_0 = 6000 \text{ (Mpc } h^{-1})^3$  for the quasars and  $10\,000 \text{ (Mpc } h^{-1})^3$  for the LRGs. These values match the amplitude of the power spectrum at  $k \sim 0.15 h \text{ Mpc}^{-1}$ , which is the optimal choice for BAO analyses (Font-Ribera et al. 2014a).

## 5.5 Weights for imaging systematics

We follow a similar approach to previous BOSS and eBOSS studies (Ross et al. 2012, 2017; Ata et al. 2018; Bautista et al. 2018) and determine weights to correct for trends with properties of the imaging and Galactic foregrounds based on linear regression. The method is most similar to Bautista et al. (2018). A multivariate linear regression is performed, comparing HEALPIX (Gorski et al. 2005) maps of the projected sample density to those of imaging properties and Galactic foregrounds. The imaging conditions of SDSS are mapped

<sup>23</sup>The overall variation in the  $n(z)$  is accounted for with the weights determined in Section 5.5



**Figure 8.** The number density of LRGs (top) and quasars (bottom) as a function of redshift, for the NGC and SGC. The vertical lines display the redshift cuts applied to create the clustering catalogues.

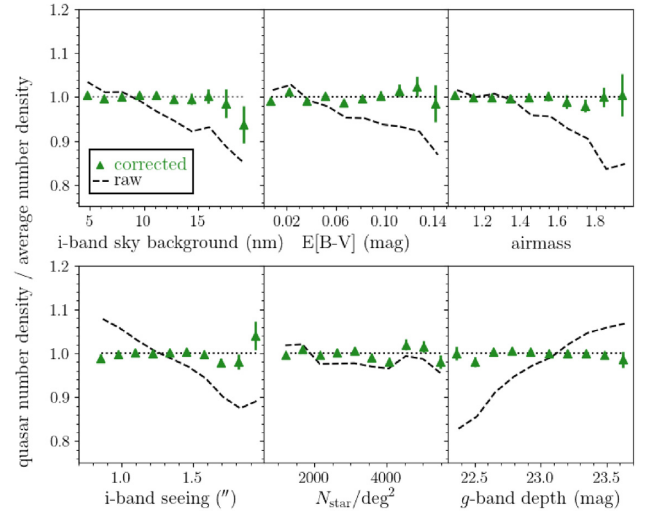
at a HEALPIX resolution<sup>24</sup>  $N_{\text{side}} = 512$ . The map was created from a dense random sample that directly queried the SDSS imaging properties over the full eBOSS footprint. We will use maps of the depth in the  $g$  band, the PSF size in the  $i$  band, the sky background in the  $i$  band, the airmass, and the Schlegel, Finkbeiner & Davis (1998) Galactic extinction ( $E[B - V]$ ). The particular choice of band is mostly arbitrary, as the SDSS imaging properties are highly correlated between bands. We use the same SDSS stellar density map, at  $N_{\text{side}} = 256$  as used in previous analyses (e.g. Ata et al. 2018; Bautista et al. 2018).

We fit for a different set of maps for the LRGs and quasars. For each the spectroscopic completeness ( $w_{\text{cp}} w_{\text{noz}}$ ) and  $w_{\text{FKP}}$  weights are applied along with the completeness cuts described in Section 5.4 to create the maps. Further, the regression for each is performed separately for the NGC and SGC and the catalogues are cut to their target redshift range. For the LRGs, this is  $0.6 < z < 1.0$  and for the quasars this is  $0.8 < z < 2.2$ . We regress against a given set of maps and, similar to the correction for redshift inefficiencies, we simply use the inverse of the fit as the weight ( $w_{\text{sys}}$ ) to apply to each object

<sup>24</sup>HEALPIX splits the sky into  $12N_{\text{side}}^2$  equal area pixels.

**Table 5.** Coefficients for the linear regressions used to determine the values of weights to correct for systematic trends with characteristics of the imaging data. The regressions are performed separately for the NGC and SGC data. Both LRGs and quasars are regressed against Galactic extinction ( $E[B - V]$ ),  $i$ -band sky background ( $\text{sky}_i$ ) and the PSF size in the  $i$ -band ( $\text{PSF}_i$ ). The extinction-corrected  $g$ -band depth ( $\text{depth}_g$ ) is additionally included in the regression against quasar density. For the LRG regression, stellar density ( $\delta_{\text{star}}$ ) is included.  $A_{\text{sys}}$  is the constant in the linear regression.

Sample	$\delta_{\text{star}}$	$\text{depth}_g$	$E[B - V]$	$\text{sky}_i$	$\text{PSF}_i$	$A_{\text{sys}}$
NGC quasars	–	0.11	–0.14	–0.038	–0.095	0.030
SGC quasars	–	0.25	–0.12	–0.075	–0.12	–0.034
NGC LRGs	–0.25	–	–0.17	0.17	–0.062	0.056
SGC LRGs	–0.48	–	–0.042	0.096	–0.075	0.098



**Figure 9.** Fluctuations in projected quasar density as a function of various image properties and Galactic foregrounds, combining NGC and SGC results (but normalizing them separately). The dashed curves show the result before weights for  $g$ -band depth and  $E[B - V]$  are applied.

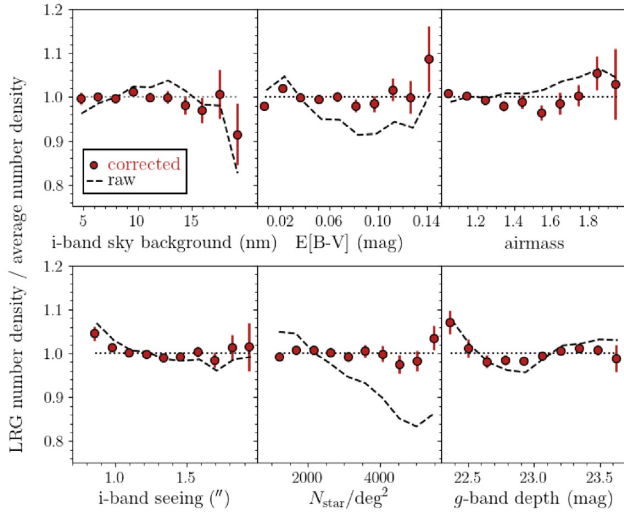
to correct for imaging systematics. We define

$$w_{\text{sys}} = [A_{\text{sys}} + \mathbf{C}_{\text{sys}} \cdot \mathbf{P}]^{-1}, \quad (13)$$

where  $\mathbf{C}_{\text{sys}}$  is the vector representing the coefficients fit to the set of maps with values  $\mathbf{P}$  at the location of a given object.

For the quasars, we use the maps that capture the SDSS imaging depth in the  $g$ -band and Galactic extinction, as was done in the DR14 analysis (Ata et al. 2018). We further include maps of the sky background and seeing. The coefficients determined from the regressions are included in Table 5. Fig. 9 displays fluctuations in the projected quasar density as a function of the imaging properties and Galactic foregrounds considered in our analysis. We have combined NGC and SGC for these results, but normalized them separately. (Given that the five SDSS imaging bands were observed nearly simultaneously via drift scan, the specific bands are nearly perfectly correlated.) One can observe that strong trends with all maps are greatly reduced after the weights are applied. For instance, a strong trend with airmass is removed, despite us not including that map in the regression. This is due to the fact that airmass is one contributing factor to the depth. The  $\chi^2$  for the null test after applying the systematic weights (ignoring any covariance) are 6, 34, 5, 9, 20, 5 left-to-right, top-to-bottom. Four maps were used in the regression and 10 measurement bins are used in this test, so the total  $\chi^2/\text{dof}$  is





**Figure 10.** Fluctuations in projected LRG density as a function of various image properties and Galactic foregrounds, combining NGC and SGC results (but normalizing them separately). The dashed curves show the result before weights for stellar density and  $E(B - V)$  are applied.

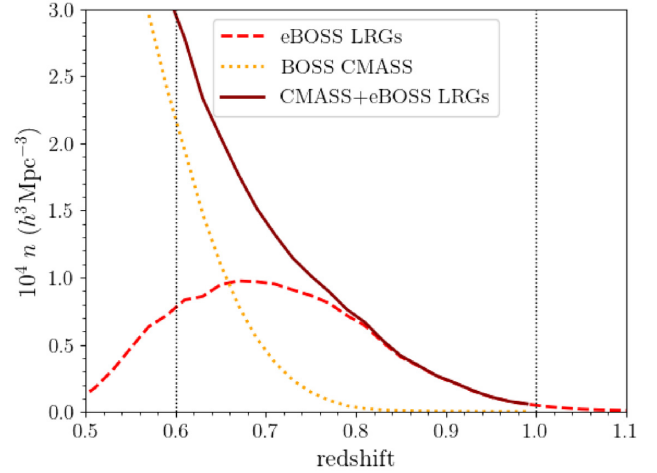
79/55. The strongest residual is for the Galactic extinction ( $E[B - V]$ ), despite the fact it was one of the maps used in the regression. The impact of any residual uncertainty on RSD or BAO analyses with respect to these imaging weights is studied further in Hou et al. (2020) and Neveux et al. (2020). Primordial non-Gaussianity studies are particularly sensitive to the large-scale power that can be introduced by these kind of systematic fluctuations (Huterer, Cunha & Fang 2013; Ross et al. 2013). The impact of residuals and whether further cleaning for such primordial non-Gaussianity analyses is possible is being investigated by Rezaie et al. (in preparation), Mueller et al. (in preparation).

For the LRGs, the systematic correlation is strongest with stellar density (consistent with Bautista et al. 2018). We additionally include the Galactic extinction, sky background, and seeing maps in the regression. The coefficients determined from the regressions are included in Table 5. Fig. 10 displays fluctuations in the LRG density before (dashed curves) and after (points with error-bars) the weights are applied. The  $\chi^2$  for the null test after applying the systematic weights (ignoring any covariance) are 6, 30, 14, 13, 8, 20 left-to-right, top-to-bottom. The total  $\chi^2/\text{dof}$  is 91/55. Similar to the results for the quasars, the strongest residual is for the Galactic extinction ( $E[B - V]$ ), despite the fact it was one of the maps used in the regression. The impact of any residual uncertainty on RSD or BAO analyses with respect to these imaging weights is studied further in Gil-Marín et al. (2020) and Bautista et al. (2020).

### 5.6 Assigning radial selection function to randoms

In order to assign redshifts and relevant information to the random catalogues, we follow the same procedure as in previous BOSS and eBOSS analyses and randomly select redshifts from the relevant observed sample. A difference for the LRG and quasar catalogues, however, is that all of the columns that are used for the data sample are also used for the random catalogue.

As a first step, the  $C_{\text{eBOSS}}$  information is copied to the  $w_{\text{sys}}$  column of the random catalogue. Then, for each RA, Dec in the random catalogue, a random LRG/quasar is selected. Its redshift,  $w_{\text{cp}}$ ,  $w_{\text{noz}}$ ,



**Figure 11.** The number density of BOSS CMASS, eBOSS LRGs, and their combination, in the NGC region. The combined sample is used within the eBOSS footprint, while CMASS only is used outside of the eBOSS footprint.

and  $w_{\text{FKP}}$  are assigned to the row. Its  $w_{\text{sys}}$  is multiplied by the existing  $w_{\text{sys}}$  to provide the total value for this quantity.

Treating the randoms in this fashion means that the random catalogues are processed in exactly the same way as the data file in order to calculate any statistics. Namely, the total contribution for any data/random point is given by

$$w_{\text{tot}} = w_{\text{sys}} w_{\text{cp}} w_{\text{noz}}. \quad (14)$$

We further recommend multiplying both the data and random  $w_{\text{tot}}$  by  $w_{\text{FKP}}$  in order to produce more optimally weighted clustering statistics.

### 5.7 Combining eBOSS LRGs and BOSS CMASS

We combine eBOSS LRGs and BOSS CMASS galaxies with  $z > 0.6$  in order to create one sample to be used for cosmological analysis. For CMASS, we make few alterations to the sample defined in Reid et al. (2016). We first cut the sample to  $z > 0.6$ . We then enforce that the ratio of weighted randoms to weighted data is the same for both CMASS and eBOSS LRGs. We then determine which CMASS galaxies are within the eBOSS footprint, using MANGLE to match to the eBOSS sectors. We assume all eBOSS LRGs are within the CMASS footprint. Within the eBOSS sectors, the  $n(z)$  is recalculated by adding the LRG and CMASS  $n(z)$ . The  $n(z)$  for each sample is shown in Fig. 11. This new  $n(z)$  is used to recalculate the  $w_{\text{FKP}}$  within the eBOSS region. Outside of the eBOSS region, the CMASS catalogues remain the same. The definition of the spectroscopic completion weights is different in CMASS. Thus, for convenience, we produce a  $w_{\text{tot}}$  column,<sup>25</sup> applying the appropriate algorithm to each sample.

Some statistics for the combined LRG+CMASS sample are given in Table 6. Overall, BOSS CMASS galaxies make up slightly more than half of the total sample and the area they occupy is more than twice that of eBOSS LRGs. One can see the footprints of the eBOSS and CMASS areas in Fig. 1. The majority (65 per cent) of the CMASS SGC was observed by the eBOSS LRG program, while eBOSS covered 37 per cent of the NGC CMASS area. Thus, the fraction of the sample that is comprised of eBOSS LRGs is different in

<sup>25</sup>It is named ‘WEIGHT\_ALL\_NOFKP’.



**Table 6.** Some statistics for the area and number of eBOSS and BOSS galaxies that enter the combined LRG+CMASS sample with  $0.6 < z < 1.0$ . For each division of the sample, we quote the  $N_{\text{eff}}$ , which is the sum of number of galaxies weighted by the close-pair and redshift failure weights.

	BOSS only			Overlap			Combined		
	SGC	NGC	Total	SGC	NGC	Total	SGC	NGC	Total
Area (deg <sup>2</sup> )	884	4368	5251	1676	2566	4242	2560	6934	9493
$N_{\text{eff}}$ eBOSS	0	0	0	71 427	113 868	185 295	71 427	113 868	185 295
$N_z$ eBOSS	0	0	0	67 316	107 500	174 816	67 316	107 500	174 816
$N_{\text{eff}}$ BOSS	16 645	95 247	111 892	41 753	63 112	104 865	58 398	158 358	216 756
$N_z$ BOSS	15 495	88 952	104 447	38 906	59 289	98 195	54 401	148 241	202 642
$N_{\text{eff}}$ BOSS+eBOSS	16 645	95 247	111 892	113 180	176 980	290 160	129 825	272 226	402 052
$N_z$ BOSS+eBOSS	15 495	88 952	104 447	106 222	166 789	273 011	121 717	255 741	377 458

each region, 55 compared to 43 per cent. The projected angular number density of galaxies with  $0.6 < z < 1.0$  is nearly twice as high for the eBOSS LRGs compared to CMASS (44 deg<sup>-2</sup> compared to 23 deg<sup>-2</sup>).

The combined LRG+CMASS sample is run through the reconstruction algorithm described in Burden et al. (2014) and Bautista et al. (2018). This moves overdensities back along an estimate of the vector of linear motion, removing some of the non-linear dispersion signal from the data. This, in turn, sharpens the BAO peak and thus increases the precision of BAO-based distance-redshift measurement (Eisenstein et al. 2007). In our companion papers, reconstruction is only applied for BAO measurements and not those that use the RSD signal. The LRG catalogue differs from the quasar catalogue in this regard, as reconstruction was not applied to the quasars due to the lower number density. We provide LRG catalogues with and without reconstruction applied. Further details on how the reconstruction algorithm was applied are presented in Bautista et al. (2020).

## 6 CONCLUSIONS

We have described the creation of LSS catalogues for the eBOSS DR16 LRG and quasar samples. The LRG catalogue is combined with BOSS CMASS and the combined catalogue contains 377 458 galaxies with  $0.6 < z < 1.0$  intended for cosmological analysis. Likewise, our catalogue contains 343 708 quasars with  $0.8 < z < 2.2$ . For each sample, there is a random sample that is at least 40 times more dense and approximates the respective 3D (RA, Dec, redshift) selection function. Weights are provided for both the data and random samples in order to ensure the randoms do match the selection function and optimize the signal to noise of the clustering measurements. These catalogues are available to the public at <https://data.sdss.org/sas/dr16/eboss/lss/catalogs/DR16/>.

Our descriptions allow our results to be reproduced. They further allow the companion analyses to study systematic uncertainties imparted during any part of the process. Thus, in addition to analysing the two-point statistics of the catalogues reported here, companion papers use this information to simulate the LRG and quasar samples and demonstrate that systematic uncertainties are a sub-dominant component to the LRG and quasar cosmological results. In particular:

(i) The uncertainties on quasar redshifts are characterized in Section 4.2, based on inputs from Lyke (2020) and including the rate of catastrophic failures. In Smith et al. (2020), these results are used to create simulations with varying assumptions on the redshift error distribution. Neveux et al. (2020) and Hou et al. (2020) apply their analyses to these simulations in order to quantify the level of systematic uncertainty introduced from these redshift uncertainties.

(ii) The completeness map and  $n(z)$  (Section 5.4) are used in Zhao et al. (2020a) to produce mock LRG and quasar surveys that are used for covariance matrix estimation.

(iii) Zhao et al. (2020a) further approximate the process used to determine weights for spectroscopic completeness (described in Section 5.3) and imaging systematics (described in Section 5.5). This allows for the estimation of systematic uncertainty related to these corrections. The results for LRG clustering measurements are in Bautista et al. (2020), Gil-Marín et al. (2020), and for quasar clustering measurements in Neveux et al. (2020) and Hou et al. (2020). In all cases, the total observational systematic uncertainty is found to be subdominant compared to the statistical uncertainty.

In addition to the LRG and LSS quasar analyses, eBOSS DR16 includes studies of ELG clustering and the Ly $\alpha$  forest. The ELG catalogues are presented in Raichoor et al. (2020), and analysed in Tamone et al. (2020), de Mattia et al. (2020), and Alam et al. (2020b). The DR16 Ly $\alpha$  sample is presented and analysed in du Mas des Bourboux et al. (2020). The results from all of the eBOSS tracers are used in Alam et al. (2020a) in order to update our understanding of cosmology.

The public release of these catalogues marks the end of the eBOSS experiment. After two decades of cosmology, eBOSS represents the conclusion of LSS surveys performed by the Sloan telescope. DESI (Aghamousa et al. 2016a,b) is the spectroscopic cosmology program that is the natural successor to eBOSS. DESI achieved first light in late 2019 and will have approximately 20 times the power of the Sloan telescope + BOSS spectrograph for conducting galaxy surveys. We expect significant research will be required to update the methods for catalogue creation presented here, in particular to model focal plane incompleteness and  $n(z)$  variation, in order to maintain systematic uncertainties that are below the sub-percent statistical precision that is expected from that program.

## ACKNOWLEDGEMENTS

AJR is grateful for support from the Ohio State University Center for Cosmology and Particle Physics. SA is supported by the European Research Council through the COSFORM Research Grant (#670193). BL and ADM were supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under Award Number DE-SC0019022. Authors acknowledge support from the ANR eBOSS project (ANR-16-CE31-0021) of the French National Research Agency. GR is supported by the National Research Foundation of Korea (NRF) through Grants No. 2017R1E1A1A01077508 and No. 2020R1A2C1005655 funded by the Korean Ministry of Education, Science and Technology (MoEST), and by the faculty research fund of Sejong University.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is [www.sdss.org](http://www.sdss.org). This work made use of the facilities and staff of the UK Sciama High Performance Computing cluster supported by the ICG, SEPNet and the University of Portsmouth. In addition, this research relied on resources provided to the eBOSS Collaboration by the National Energy Research Scientific Computing Center (NERSC). NERSC is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

## DATA AVAILABILITY

The data described in this paper are available to the public at <https://data.sdss.org/sas/dr16/eboass/lss/catalogs/DR16/> and are described at <https://www.sdss.org/dr16/spectro/lss/>.

## REFERENCES

- Abolfathi B. et al., 2018, *ApJS*, 235, 42  
 Aghamousa A. et al., 2016a, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))  
 Aghamousa A. et al., 2016b, preprint ([arXiv:1611.00037](https://arxiv.org/abs/1611.00037))  
 Ahumada R. et al., 2020, *ApJS*, 249, 3  
 Alam S. et al., 2020b, preprint ([arXiv:2007.09004](https://arxiv.org/abs/2007.09004))  
 Alam S., Ata M., Bailey S. et al., 2017, *MNRAS*, 470, 2617  
 Alam S., Peacock J. A., Kraljic K., Ross A. J., Comparat J., 2020, *MNRAS*, 497, 581  
 Alam S. et al., 2020a, preprint ([arXiv:2007.08991](https://arxiv.org/abs/2007.08991))  
 Albareti F. D. et al., 2017, *ApJS*, 233, 25  
 Allende Prieto C., Koesterke L., Hubeny I., Bautista M. A., Barklem P. S., Nahar S. N., 2018, *A&A*, 618, A25  
 Anderson L. et al., 2012, *MNRAS*, 427, 3435  
 Ata M. et al., 2018, *MNRAS*, 473, 4773  
 Bautista J. E. et al., 2017, *A&A*, 603, A12  
 Bautista J. E. et al., 2018, *ApJ*, 863, 110  
 Bautista J. E. et al., 2020, preprint ([arXiv:2007.08993](https://arxiv.org/abs/2007.08993))  
 Bianchi D., Percival W. J., 2017, *MNRAS*, 472, 1106  
 Blanton M. et al., 2003, *AJ*, 125, 2276  
 Blanton M. R. et al., 2005, *AJ*, 129, 2562  
 Blanton M. R. et al., 2017, *AJ*, 154, 28  
 Blomqvist M. et al., 2019, *A&A*, 629, A86  
 Bolton A. S. et al., 2012, *AJ*, 144, 144  
 Bovy J. et al., 2012, *ApJ*, 749, 41  
 Burden A., Percival W. J., Manera M., Cuesta A. J., Magana M. V., Ho S., 2014, *MNRAS*, 445, 3152  
 Chabanier S. et al., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 017  
 Clerc N. et al., 2016, *MNRAS*, 463, 4490  
 Cole S. et al., 2005, *MNRAS*, 362, 505  
 Comparat J. et al., 2016, *MNRAS*, 461, 1076  
 Cool R. J. et al., 2013, *ApJ*, 767, 118  
 Dawson K. et al., 2013, *AJ*, 145, 10  
 Dawson K. S. et al., 2016, *AJ*, 151, 44  
 de Mattia A. et al., 2020, preprint ([arXiv:2009.09008](https://arxiv.org/abs/2009.09008))  
 de Sainte Agathe V. et al., 2019, *A&A*, 629, A85  
 Dey A. et al., 2019, *AJ*, 157, 168  
 Doi M. et al., 2010, *AJ*, 139, 1628  
 du Mas des Bourboux H. et al., 2020, preprint ([arXiv:2007.08995](https://arxiv.org/abs/2007.08995))  
 Dwelly T. et al., 2017, *MNRAS*, 469, 1065  
 Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605  
 Eisenstein D. J. et al., 2005, *ApJ*, 633, 560  
 Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007, *ApJ*, 664, 675  
 Eisenstein D. J. et al., 2011, *AJ*, 142  
 Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23  
 Finkbeiner D. P. et al., 2016, *ApJ*, 822, 66  
 Font-Ribera A. et al., 2014, *J. Cosmol. Astropart. Phys.*, 5, 023  
 Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748  
 Gil-Marín H. et al., 2020, preprint ([arXiv:2007.08994](https://arxiv.org/abs/2007.08994))  
 Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759  
 Gunn J. E. et al., 1998, *AJ*, 116, 3040  
 Gunn J. E. et al., 2006, *AJ*, 131, 2332  
 Hahn C., Scoccimarro R., Blanton M. R., Tinker J. L., Rodríguez-Torres S. A., 2017, *MNRAS*, 467, 1940  
 Hou J. et al., 2020, preprint ([arXiv:2007.08998](https://arxiv.org/abs/2007.08998))  
 Hutchinson T. A. et al., 2016, *AJ*, 152, 205  
 Huterer D., Cunha C. E., Fang W., 2013, *MNRAS*, 432, 2945  
 Jensen T. W. et al., 2016, *ApJ*, 833, 199  
 Kaiser N., 1987, *MNRAS*, 227, 1  
 Kaiser N. et al., 2010, in Stepp L. M., Gilmozzi R., Hall H. J., eds, *Proc. SPIE Conf. Ser. Vol. 7733, Ground-based and Airborne Telescopes III*. SPIE, Bellingham, p. 77330E  
 Lang D., 2014, *AJ*, 147, 108  
 Lang D., Hogg D. W., Schlegel D. J., 2016, *AJ*, 151, 36  
 Law N. M. et al., 2009, *PASP*, 121, 1395  
 Lin S., 2020, preprint ([arXiv:2007.08996](https://arxiv.org/abs/2007.08996))  
 Lyke B., 2020, *ApJS*, 250, 8  
 Meisner A. M., Lang D., Schlafly E. F., Schlegel D. J., 2019, *PASP*, 131, 124504  
 Mohammad F. et al., 2020, *MNRAS*, 498, 128  
 Morganson E. et al., 2015, *ApJ*, 806, 244  
 Myers A. D. et al., 2015, *ApJS*, 221, 27  
 Neveux N. et al., 2020, preprint ([arXiv:2007.08999](https://arxiv.org/abs/2007.08999))  
 Padmanabhan N. et al., 2008, *ApJ*, 674, 1217  
 Palanque-Delabrouille N. et al., 2016, *A&A*, 587, A41  
 Pâris I. et al., 2018, *A&A*, 613, A51  
 Percival W. J. et al., 2010, *MNRAS*, 401, 2148  
 Prakash A. et al., 2016, *ApJS*, 224, 34  
 Raichoor A. et al., 2017, *MNRAS*, 471, 3955  
 Raichoor A. et al., 2020, preprint ([arXiv:2007.09007](https://arxiv.org/abs/2007.09007))  
 Rau A. et al., 2009, *PASP*, 121, 1334  
 Reid B. et al., 2016, *MNRAS*, 455, 1553  
 Ross A. J. et al., 2012, *MNRAS*, 424, 564  
 Ross A. J. et al., 2013, *MNRAS*, 428, 1116  
 Ross A. J. et al., 2017, *MNRAS*, 464, 1168  
 Rossi G. et al., 2020, preprint ([arXiv:2007.09002](https://arxiv.org/abs/2007.09002))  
 Schlafly E. F. et al., 2012, *ApJ*, 756, 158

- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- Shen Y. et al., 2015, *ApJS*, 216, 4
- Smee S. A. et al., 2013, *AJ*, 146, 32
- Smith A. et al., 2020, preprint ([arXiv:2007.09003](https://arxiv.org/abs/2007.09003))
- Smith J. A. et al., 2002, *AJ*, 123, 2121
- Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, *MNRAS*, 387, 1391
- Tamone A. et al., 2020, preprint ([arXiv:2007.09009](https://arxiv.org/abs/2007.09009))
- Wang Y. et al., 2020, preprint ([arXiv:2007.09010](https://arxiv.org/abs/2007.09010))
- Wright E. L. et al., 2010, *AJ*, 140, 1868
- York D. G. et al., 2000, *AJ*, 120, 1579
- Zhao C. et al., 2020, preprint ([arXiv:2007.08997](https://arxiv.org/abs/2007.08997))
- Zhao G. et al., 2020, preprint ([arXiv:2007.09011](https://arxiv.org/abs/2007.09011))
- Zhu G., 2016, preprint ([arXiv:1606.07156](https://arxiv.org/abs/1606.07156))
- <sup>1</sup>Center for Cosmology and Astro-Particle Physics, Ohio State University, Columbus, Ohio 43210, USA
- <sup>2</sup>Institute of Cosmology and Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth PO1 3FX, UK
- <sup>3</sup>School of Physics and Astronomy, University of St Andrews, St Andrews KY16 9SS, UK
- <sup>4</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
- <sup>5</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
- <sup>6</sup>IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
- <sup>7</sup>Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstr., D-85748 Garching bei München, Germany
- <sup>8</sup>Department Physics and Astronomy, University of Utah, 115 S 1400 E, Salt Lake City, UT 84112, USA
- <sup>9</sup>Institut de Ciències del Cosmos, Universitat de Barcelona, ICCUB, Martí i Franquès 1, E-08028 Barcelona, Spain
- <sup>10</sup>Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain
- <sup>11</sup>Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA
- <sup>12</sup>Waterloo Centre for Astrophysics, Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada
- <sup>13</sup>Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada
- <sup>14</sup>Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, NY 12211, USA
- <sup>15</sup>Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada
- <sup>16</sup>Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland
- <sup>17</sup>Department of Physics and Astronomy, Ohio University, 251B Clippinger Labs, Athens, OH 45701, USA
- <sup>18</sup>Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA
- <sup>19</sup>Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK
- <sup>20</sup>National Astronomy Observatories, Chinese Academy of Science, Beijing 100101, P.R. China
- <sup>21</sup>School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, P. R. China
- <sup>22</sup>Apache Point Observatory and New Mexico State University, PO Box 59, Sunspot, NM 88349, USA
- <sup>23</sup>Centro de Investigaciones Energeticas, Medioambientales y Tecnológicas (CIEMAT), Madrid 28040, Spain
- <sup>24</sup>Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea
- <sup>25</sup>Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA
- <sup>26</sup>Instituto de Física, Universidad Nacional Autónoma de México, Apdo. Postal 20-364 Ciudad de México, México
- <sup>27</sup>Aix Marseille Université, CNRS, CNES, LAM, Marseille, 13007, France
- <sup>28</sup>CPPM, Aix Marseille Université, CNRS/IN2P3, Marseille, 13007, France
- <sup>29</sup>Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Av. Universidad s/n, 62210 Cuernavaca, Mor., México
- <sup>30</sup>PITT PACC, Department of Physics and Astronomy, University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260, USA
- <sup>31</sup>Centro de Astronomía (CITEVA), Universidad de Antofagasta, Avenida Angamos 601, Antofagasta 1270300, Chile
- <sup>32</sup>Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA
- <sup>33</sup>Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
- <sup>34</sup>Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
- <sup>35</sup>Indian Institute of Astrophysics, Koramangala, Bangalore 560034, India

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.